Simon & Schuster Data Challenge

March 10, 2020

```python
[1]: import pandas as pd
     import numpy as np
     import seaborn as sns
     import matplotlib.pyplot as plt
     %matplotlib inline
```

```
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\tools\_testing.py:19:
FutureWarning: pandas.util.testing is deprecated. Use the functions in the
public API at pandas.testing instead.
  import pandas.util.testing as tm
```

```python
[2]: data = pd.read_csv("C:/Users/sonje/Desktop/simon & schuster/sample_data.
     →csv",parse_dates = True)
     pd.set_option('display.max_columns', None)
     pd.set_option('display.max_rows', None)
```

```python
[3]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6539 entries, 0 to 6538
Columns: 106 entries, entity to appledistributor
dtypes: float64(4), int64(91), object(11)
memory usage: 5.3+ MB
```

```python
[4]: data.describe()
```

```
[4]:            fromdate    titlecount        todate  amazontotalreviews  \
       count  6.539000e+03  6539.000000  6.539000e+03         6522.000000
       mean   2.018078e+07     1.599174  2.018080e+07          242.896044
       std    7.319859e+03     0.748654  7.321285e+03           57.042772
       min    2.017040e+07     1.000000  2.017040e+07          120.000000
       25%    2.017121e+07     1.000000  2.017121e+07          200.000000
       50%    2.018071e+07     1.000000  2.018071e+07          250.000000
       75%    2.019013e+07     2.000000  2.019013e+07          283.000000
       max    2.019083e+07     4.000000  2.019083e+07          331.000000

              minappleprice  amazontotalrevenue  accruedpreordersoldunits  \
```

count	6539.000000	6539.000000	6539.000000
mean	118.859459	18074.906866	0.235204
std	323.464108	30999.078107	2.776806
min	0.000000	0.000000	0.000000
25%	0.000000	1998.000000	0.000000
50%	0.000000	7191.000000	0.000000
75%	0.000000	17982.000000	0.000000
max	999.000000	571692.000000	82.000000

	amazonlumpedpreordersubscriptionrevenue	preordersalesrevenue \
count	6539.0	6539.000000
mean	0.0	155.503288
std	0.0	8769.366153
min	0.0	0.000000
25%	0.0	0.000000
50%	0.0	0.000000
75%	0.0	0.000000
max	0.0	501420.000000

	foundatamazon	barnesandnobleaccruedpreorderrevenue \
count	6539.000000	6539.0
mean	0.998165	0.0
std	0.042803	0.0
min	0.000000	0.0
25%	1.000000	0.0
50%	1.000000	0.0
75%	1.000000	0.0
max	1.000000	0.0

	amazonpreordersubscriptionunits	amazonlumpedpreordersalesrevenue \
count	6539.0	6539.000000
mean	0.0	155.503288
std	0.0	8769.366153
min	0.0	0.000000
25%	0.0	0.000000
50%	0.0	0.000000
75%	0.0	0.000000
max	0.0	501420.000000

	foundatbarnesandnoble	maxprice	lumpedpreordersubscriptionunits \
count	6539.000000	6539.000000	6539.0
mean	0.265331	1120.283835	0.0
std	0.441543	480.250221	0.0
min	0.000000	0.000000	0.0
25%	0.000000	799.000000	0.0
50%	0.000000	999.000000	0.0
75%	1.000000	1595.000000	0.0

```
max                1.000000  3499.000000                                0.0

        applepreorderrevenue  amazontotalunits   totalrevenue  \
count                 6539.0       6539.000000    6539.000000
mean                     0.0         19.246674   19047.171127
std                      0.0         36.567037   32704.552550
min                      0.0          0.000000       0.000000
25%                      0.0          2.000000    1998.000000
50%                      0.0          7.000000    7675.000000
75%                      0.0         16.000000   20035.000000
max                      0.0        781.000000  571692.000000

        barnesandnobleaccruedpreorderunits   preorderunits  \
count                               6539.0    6539.000000
mean                                   0.0       0.210124
std                                    0.0      11.978933
min                                    0.0       0.000000
25%                                    0.0       0.000000
50%                                    0.0       0.000000
75%                                    0.0       0.000000
max                                    0.0     685.000000

        amazonsubscriptionrevenue  amazonlumpedpreorderrevenue  \
count                 6539.000000                  6539.000000
mean                  2363.101392                   155.503288
std                   7569.937102                  8769.366153
min                      0.000000                     0.000000
25%                      0.000000                     0.000000
50%                      0.000000                     0.000000
75%                      0.000000                     0.000000
max                 216775.000000                501420.000000

        maxbarnesandnoblelistprice  maxlistprice  barnesandnoblesoldunits  \
count                       6539.0  6539.000000              6539.000000
mean                           0.0   813.977672                 0.002447
std                            0.0   716.871551                 0.139899
min                            0.0     0.000000                 0.000000
25%                            0.0     0.000000                 0.000000
50%                            0.0   799.000000                 0.000000
75%                            0.0  1458.000000                 0.000000
max                            0.0  3499.000000                 8.000000

        maxapplelistprice  amazonaccruedpreordersubscriptionrevenue  \
count              6539.0                                    6539.0
mean                  0.0                                       0.0
std                   0.0                                       0.0
min                   0.0                                       0.0
```

25%	0.0	0.0
50%	0.0	0.0
75%	0.0	0.0
max	0.0	0.0

	accruedpreordersubscriptionunits	amazonkenpread	applesoldunits \
count	6539.0	6539.0	6539.000000
mean	0.0	0.0	0.970791
std	0.0	0.0	10.904038
min	0.0	0.0	0.000000
25%	0.0	0.0	0.000000
50%	0.0	0.0	0.000000
75%	0.0	0.0	0.000000
max	0.0	0.0	451.000000

	minamazonlistprice	minprice	bookscanreported \
count	6539.000000	6539.000000	6539.000000
mean	754.291329	1070.648876	0.506194
std	706.126853	512.288446	0.500000
min	0.000000	0.000000	0.000000
25%	0.000000	732.000000	0.000000
50%	732.000000	999.000000	1.000000
75%	1357.000000	1579.000000	1.000000
max	3499.000000	3499.000000	1.000000

	appleaccruedpreorderunits	preorderperiod	lumpedpreordersoldunits \
count	6539.0	6539.000000	6539.000000
mean	0.0	0.022328	0.210124
std	0.0	0.147758	11.978933
min	0.0	0.000000	0.000000
25%	0.0	0.000000	0.000000
50%	0.0	0.000000	0.000000
75%	0.0	0.000000	0.000000
max	0.0	1.000000	685.000000

	accruedpreorderrevenue	maxamazonlistprice \
count	6539.000000	6539.000000
mean	182.564918	813.977672
std	2329.724787	716.871551
min	0.000000	0.000000
25%	0.000000	0.000000
50%	0.000000	799.000000
75%	0.000000	1458.000000
max	88478.000000	3499.000000

	amazonaccruedpreorderrevenue	barnesandnoblepreorderunits \
count	6539.000000	6539.0

		182.564918		0.0
mean		182.564918		0.0
std		2329.724787		0.0
min		0.000000		0.0
25%		0.000000		0.0
50%		0.000000		0.0
75%		0.000000		0.0
max		88478.000000		0.0

	amazonaccruedpreorderunits	minbarnesandnoblelistprice \
count	6539.000000	6539.0
mean	0.235204	0.0
std	2.776806	0.0
min	0.000000	0.0
25%	0.000000	0.0
50%	0.000000	0.0
75%	0.000000	0.0
max	82.000000	0.0

	appleaccruedpreorderrevenue	applelumpedpreorderunits \
count	6539.0	6539.0
mean	0.0	0.0
std	0.0	0.0
min	0.0	0.0
25%	0.0	0.0
50%	0.0	0.0
75%	0.0	0.0
max	0.0	0.0

	applesalesrevenue	preordersoldunits	lumpedpreordersalesrevenue \
count	6539.000000	6539.000000	6539.000000
mean	969.819850	0.210124	155.503288
std	10893.133995	11.978933	8769.366153
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000
75%	0.000000	0.000000	0.000000
max	450549.000000	685.000000	501420.000000

	totalunits	maxamazonprice	amazonlumpedpreordersoldunits \
count	6539.000000	6539.000000	6539.000000
mean	20.219911	1110.416425	0.210124
std	38.000744	485.727725	11.978933
min	0.000000	0.000000	0.000000
25%	2.000000	795.500000	0.000000
50%	7.000000	999.000000	0.000000
75%	18.000000	1595.000000	0.000000
max	781.000000	3499.000000	685.000000

```
       amazonaveragerating  accruedpreorderunits  amazonsalesrevenue  \
count          6523.000000           6539.000000         6539.000000
mean            453.302162              0.235204        15711.805475
std              13.736645              2.776806        29237.397462
min             410.000000              0.000000            0.000000
25%             440.000000              0.000000         1795.000000
50%             460.000000              0.000000         6392.000000
75%             470.000000              0.000000        14355.000000
max             470.000000             82.000000       571692.000000

       barnesandnoblesalesrevenue  minlistprice  \
count                 6539.000000   6539.000000
mean                     2.444410    670.734975
std                    139.759581    735.657320
min                      0.000000      0.000000
25%                      0.000000      0.000000
50%                      0.000000    637.000000
75%                      0.000000   1357.000000
max                   7992.000000   3499.000000

       amazonaccruedpreordersalesrevenue  amazonlumpedpreorderunits  \
count                        6539.000000                6539.000000
mean                          182.564918                   0.210124
std                          2329.724787                  11.978933
min                             0.000000                   0.000000
25%                             0.000000                   0.000000
50%                             0.000000                   0.000000
75%                             0.000000                   0.000000
max                         88478.000000                 685.000000

       lumpedpreordersubscriptionrevenue  applelumpedpreorderrevenue  \
count                             6539.0                      6539.0
mean                                 0.0                         0.0
std                                  0.0                         0.0
min                                  0.0                         0.0
25%                                  0.0                         0.0
50%                                  0.0                         0.0
75%                                  0.0                         0.0
max                                  0.0                         0.0

       minamazonprice  subscriptionunits  amazonpreorderunits  \
count     6539.000000        6539.000000          6539.000000
mean      1070.526686           1.580670             0.210124
std        512.448512           5.063503            11.978933
min          0.000000           0.000000             0.000000
25%        732.000000           0.000000             0.000000
```

50%	999.000000	0.000000	0.000000
75%	1579.000000	0.000000	0.000000
max	3499.000000	145.000000	685.000000

	amazonpreorderrevenue	amazonaccruedpreordersoldunits \
count	6539.000000	6539.000000
mean	155.503288	0.235204
std	8769.366153	2.776806
min	0.000000	0.000000
25%	0.000000	0.000000
50%	0.000000	0.000000
75%	0.000000	0.000000
max	501420.000000	82.000000

	amazonpreordersoldunits	salesrevenue	preordersubscriptionrevenue \
count	6539.000000	6539.000000	6539.0
mean	0.210124	16684.069735	0.0
std	11.978933	31113.761349	0.0
min	0.000000	0.000000	0.0
25%	0.000000	1795.000000	0.0
50%	0.000000	6800.000000	0.0
75%	0.000000	15950.000000	0.0
max	685.000000	571692.000000	0.0

	lumpedpreorderunits	amazonaccruedpreordersubscriptionunits \
count	6539.000000	6539.0
mean	0.210124	0.0
std	11.978933	0.0
min	0.000000	0.0
25%	0.000000	0.0
50%	0.000000	0.0
75%	0.000000	0.0
max	685.000000	0.0

	applepreorderunits	subscriptionrevenue	soldunits \
count	6539.0	6539.000000	6539.000000
mean	0.0	2363.101392	18.639241
std	0.0	7569.937102	37.757858
min	0.0	0.000000	0.000000
25%	0.0	0.000000	1.000000
50%	0.0	0.000000	7.000000
75%	0.0	0.000000	14.000000
max	0.0	216775.000000	781.000000

	minapplelistprice	barnesandnoblepreorderrevenue \
count	6539.0	6539.0
mean	0.0	0.0

std	0.0	0.0
min	0.0	0.0
25%	0.0	0.0
50%	0.0	0.0
75%	0.0	0.0
max	0.0	0.0

	amazonsubscriptionunits	amazonlumpedpreordersubscriptionunits \
count	6539.000000	6539.0
mean	1.580670	0.0
std	5.063503	0.0
min	0.000000	0.0
25%	0.000000	0.0
50%	0.000000	0.0
75%	0.000000	0.0
max	145.000000	0.0

	aapreported	accruedpreordersubscriptionrevenue	foundatapple \
count	6539.000000	6539.0	6539.000000
mean	0.984095	0.0	0.115614
std	0.125116	0.0	0.319786
min	0.000000	0.0	0.000000
25%	1.000000	0.0	0.000000
50%	1.000000	0.0	0.000000
75%	1.000000	0.0	0.000000
max	1.000000	0.0	1.000000

	barnesandnoblelumpedpreorderrevenue	barnesandnoblelumpedpreorderunits \
count	6539.0	6539.0
mean	0.0	0.0
std	0.0	0.0
min	0.0	0.0
25%	0.0	0.0
50%	0.0	0.0
75%	0.0	0.0
max	0.0	0.0

	minbarnesandnobleprice	accruedpreordersalesrevenue \
count	6539.000000	6539.000000
mean	128.287047	182.564918
std	317.132231	2329.724787
min	0.000000	0.000000
25%	0.000000	0.000000
50%	0.000000	0.000000
75%	0.000000	0.000000
max	999.000000	88478.000000

```
        maxbarnesandnobleprice    preorderrevenue    pubtrackreported  \
count            6539.000000         6539.000000         6539.000000
mean              128.287047          155.503288            0.984095
std               317.132231         8769.366153            0.125116
min                 0.000000            0.000000            0.000000
25%                 0.000000            0.000000            1.000000
50%                 0.000000            0.000000            1.000000
75%                 0.000000            0.000000            1.000000
max               999.000000       501420.000000            1.000000


        lumpedpreorderrevenue    amazonsoldunits  \
count             6539.000000        6539.000000
mean               155.503288          17.666004
std               8769.366153          36.272170
min                  0.000000           0.000000
25%                  0.000000           1.000000
50%                  0.000000           6.000000
75%                  0.000000          13.000000
max             501420.000000         781.000000


        amazonpreordersubscriptionrevenue    preordersubscriptionunits  \
count                             6539.0                       6539.0
mean                                 0.0                          0.0
std                                  0.0                          0.0
min                                  0.0                          0.0
25%                                  0.0                          0.0
50%                                  0.0                          0.0
75%                                  0.0                          0.0
max                                  0.0                          0.0


        amazonpreordersalesrevenue    maxappleprice    pagelength  \
count                  6539.000000      6539.000000   2558.000000
mean                    155.503288       118.859459    336.532447
std                    8769.366153       323.464108     42.531144
min                       0.000000         0.000000    305.000000
25%                       0.000000         0.000000    320.000000
50%                       0.000000         0.000000    321.000000
75%                       0.000000         0.000000    321.000000
max                  501420.000000       999.000000    496.000000


        pricewassetbypublisher
count                  1531.0
mean                      1.0
std                       0.0
min                       1.0
25%                       1.0
50%                       1.0
```

```
75%                           1.0
max                           1.0
```

```
[5]: data.head()
```

```
[5]:              entity formattype  fromdate  titlecount  \
     0  Alone in the dark     print  20170418           1
     1  Alone in the dark     print  20170419           1
     2  Alone in the dark     print  20170420           1
     3  Alone in the dark     print  20170421           1
     4  Alone in the dark     print  20170422           1

                                         titles     todate         author  \
     0  {'amazontotalreviews': 306, 'minappleprice': 0...  20170419  Frith Banbury
     1  {'amazontotalreviews': 306, 'minappleprice': 0...  20170420  Frith Banbury
     2  {'amazontotalreviews': 306, 'minappleprice': 0...  20170421  Frith Banbury
     3  {'amazontotalreviews': 306, 'minappleprice': 0...  20170422  Frith Banbury
     4  {'amazontotalreviews': 306, 'minappleprice': 0...  20170423  Frith Banbury

        amazontotalreviews  minappleprice  amazontotalrevenue  \
     0               306.0              0                   0
     1               306.0              0                 702
     2               306.0              0                4914
     3               306.0              0                4212
     4               306.0              0                 702

        accruedpreordersoldunits  amazonlumpedpreordersubscriptionrevenue  \
     0                         0                                        0
     1                         0                                        0
     2                         0                                        0
     3                         0                                        0
     4                         0                                        0

        preordersalesrevenue  foundatamazon  barnesandnobleaccruedpreorderrevenue  \
     0                     0              1                                     0
     1                     0              1                                     0
     2                     0              1                                     0
     3                     0              1                                     0
     4                     0              1                                     0

        amazonpreordersubscriptionunits  amazonlumpedpreordersalesrevenue  \
     0                                0                                 0
     1                                0                                 0
     2                                0                                 0
     3                                0                                 0
     4                                0                                 0

        foundatbarnesandnoble  maxprice  lumpedpreordersubscriptionunits  \
```

```
0                    0          0                          0
1                    0        702                          0
2                    0        702                          0
3                    0        702                          0
4                    0        702                          0

   applepreorderrevenue  amazontotalunits  totalrevenue  \
0                     0                 6             0
1                     0                 7           702
2                     0                 7          4914
3                     0                 6          4212
4                     0                 1           702

   barnesandnobleaccruedpreorderunits  preorderunits  \
0                                    0              0
1                                    0              0
2                                    0              0
3                                    0              0
4                                    0              0

   amazonsubscriptionrevenue  amazonlumpedpreorderrevenue  \
0                          0                            0
1                          0                            0
2                          0                            0
3                          0                            0
4                          0                            0

   maxbarnesandnoblelistprice  maxlistprice  barnesandnoblesoldunits  \
0                           0             0                        0
1                           0           702                        0
2                           0           702                        0
3                           0           702                        0
4                           0           702                        0

   maxapplelistprice  amazonaccruedpreordersubscriptionrevenue  \
0                  0                                         0
1                  0                                         0
2                  0                                         0
3                  0                                         0
4                  0                                         0

   accruedpreordersubscriptionunits  amazonkenpread  applesoldunits  \
0                                 0               0               0
1                                 0               0               0
2                                 0               0               0
3                                 0               0               0
4                                 0               0               0
```

```
       minamazonlistprice   minprice   bookscanreported   appleaccruedpreorderunits  \
0                        0          0                  1                           0
1                        0          0                  1                           0
2                      702        702                  1                           0
3                      702        702                  1                           0
4                      702        702                  1                           0

       preorderperiod datepublished   lumpedpreordersoldunits  \
0                   0    2009-05-12                          0
1                   0    2009-05-12                          0
2                   0    2009-05-12                          0
3                   0    2009-05-12                          0
4                   0    2009-05-12                          0

       accruedpreorderrevenue   maxamazonlistprice   amazonaccruedpreorderrevenue  \
0                            0                    0                              0
1                            0                  702                              0
2                            0                  702                              0
3                            0                  702                              0
4                            0                  702                              0

       barnesandnoblepreorderunits   amazonaccruedpreorderunits  \
0                                 0                            0
1                                 0                            0
2                                 0                            0
3                                 0                            0
4                                 0                            0

       minbarnesandnoblelistprice   appleaccruedpreorderrevenue  \
0                                0                             0
1                                0                             0
2                                0                             0
3                                0                             0
4                                0                             0

       applelumpedpreorderunits   applesalesrevenue   preordersoldunits  \
0                              0                   0                   0
1                              0                   0                   0
2                              0                   0                   0
3                              0                   0                   0
4                              0                   0                   0

       lumpedpreordersalesrevenue   totalunits   maxamazonprice  \
0                               0            6                0
1                               0            7              702
2                               0            7              702
```

```
                               0            6            702
3                                                           
4                               0            1            702

     amazonlumpedpreordersoldunits  amazonaveragerating publishertype  \
0                                0                410.0       bigfive
1                                0                410.0       bigfive
2                                0                410.0       bigfive
3                                0                410.0       bigfive
4                                0                410.0       bigfive

     accruedpreorderunits  amazonsalesrevenue  barnesandnoblesalesrevenue  \
0                       0                   0                           0
1                       0                 702                           0
2                       0                4914                           0
3                       0                4212                           0
4                       0                 702                           0

                                amazoncategories  minlistprice  \
0  Books > Books on CD > Mystery & Thrillers ; Bo...             0
1  Books > Books on CD > Mystery & Thrillers ; Bo...             0
2  Books > Books on CD > Mystery & Thrillers ; Bo...           702
3  Books > Books on CD > Mystery & Thrillers ; Bo...           702
4  Books > Books on CD > Mystery & Thrillers ; Bo...           702

     amazonaccruedpreordersalesrevenue  amazonlumpedpreorderunits  \
0                                    0                          0
1                                    0                          0
2                                    0                          0
3                                    0                          0
4                                    0                          0

     lumpedpreordersubscriptionrevenue  applelumpedpreorderrevenue  \
0                                    0                           0
1                                    0                           0
2                                    0                           0
3                                    0                           0
4                                    0                           0

     minamazonprice  subscriptionunits  amazonpreorderunits  publisherparent  \
0                 0                  0                    0  Simon & Schuster
1                 0                  0                    0  Simon & Schuster
2               702                  0                    0  Simon & Schuster
3               702                  0                    0  Simon & Schuster
4               702                  0                    0  Simon & Schuster

     amazonpreorderrevenue  amazonaccruedpreordersoldunits  \
0                        0                               0
```

1	0	0
2	0	0
3	0	0
4	0	0

	amazonpreordersoldunits	salesrevenue	preordersubscriptionrevenue
0	0	0	0
1	0	702	0
2	0	4914	0
3	0	4212	0
4	0	702	0

	lumpedpreorderunits	amazonaccruedpreordersubscriptionunits	formats
0	0	0	Audio CD
1	0	0	Audio CD
2	0	0	Audio CD
3	0	0	Audio CD
4	0	0	Audio CD

	applepreorderunits	subscriptionrevenue	soldunits	minapplelistprice
0	0	0	6	0
1	0	0	7	0
2	0	0	7	0
3	0	0	6	0
4	0	0	1	0

	barnesandnoblepreorderrevenue	amazonsubscriptionunits
0	0	0
1	0	0
2	0	0
3	0	0
4	0	0

	amazonlumpedpreordersubscriptionunits	aapreported
0	0	1
1	0	1
2	0	1
3	0	1
4	0	1

	accruedpreordersubscriptionrevenue	foundatapple
0	0	0
1	0	0
2	0	0
3	0	0
4	0	0

```
   barnesandnoblelumpedpreorderrevenue  barnesandnoblelumpedpreorderunits  \
0                                    0                                  0
1                                    0                                  0
2                                    0                                  0
3                                    0                                  0
4                                    0                                  0


   minbarnesandnobleprice  accruedpreordersalesrevenue  \
0                       0                            0
1                       0                            0
2                       0                            0
3                       0                            0
4                       0                            0


   maxbarnesandnobleprice  preorderrevenue  pubtrackreported  \
0                       0                0                 1
1                       0                0                 1
2                       0                0                 1
3                       0                0                 1
4                       0                0                 1


   lumpedpreorderrevenue  amazonsoldunits  amazonpreordersubscriptionrevenue  \
0                      0                6                                  0
1                      0                7                                  0
2                      0                7                                  0
3                      0                6                                  0
4                      0                1                                  0


   preordersubscriptionunits  amazonpreordersalesrevenue  maxappleprice  \
0                          0                           0              0
1                          0                           0              0
2                          0                           0              0
3                          0                           0              0
4                          0                           0              0


   pagelength  pricewassetbypublisher barnesandnobledistributor  \
0         NaN                     NaN                       NaN
1         NaN                     NaN                       NaN
2         NaN                     NaN                       NaN
3         NaN                     NaN                       NaN
4         NaN                     NaN                       NaN


   appledistributor
0               NaN
1               NaN
2               NaN
3               NaN
```

```
    4              NaN
```

```
[6]: data.tail()
```

```
[6]:                  entity formattype  fromdate  titlecount  \
     6534  The Walking Dead      audio  20190822           1
     6535  The Walking Dead      audio  20190823           1
     6536  The Walking Dead      audio  20190824           1
     6537  The Walking Dead      audio  20190825           1
     6538  The Walking Dead      audio  20190826           1


                                                     titles    todate  \
     6534  {'amazontotalreviews': 275, 'minappleprice': 0...  20190823
     6535  {'amazontotalreviews': 275, 'minappleprice': 0...  20190824
     6536  {'amazontotalreviews': 275, 'minappleprice': 0...  20190825
     6537  {'amazontotalreviews': 275, 'minappleprice': 0...  20190826
     6538  {'amazontotalreviews': 275, 'minappleprice': 0...  20190827


                   author  amazontotalreviews  minappleprice  amazontotalrevenue  \
     6534  Matthew Murdock               275.0              0               81835
     6535  Matthew Murdock               275.0              0               18460
     6536  Matthew Murdock               275.0              0               38623
     6537  Matthew Murdock               275.0              0               80340
     6538  Matthew Murdock               275.0              0               43212


           accruedpreordersoldunits  amazonlumpedpreordersubscriptionrevenue  \
     6534                          0                                        0
     6535                          0                                        0
     6536                          0                                        0
     6537                          0                                        0
     6538                          0                                        0


           preordersalesrevenue  foundatamazon  \
     6534                      0              1
     6535                      0              1
     6536                      0              1
     6537                      0              1
     6538                      0              1


           barnesandnobleaccruedpreorderrevenue  amazonpreordersubscriptionunits  \
     6534                                      0                                0
     6535                                      0                                0
     6536                                      0                                0
     6537                                      0                                0
     6538                                      0                                0


           amazonlumpedpreordersalesrevenue  foundatbarnesandnoble  maxprice  \
     6534                                  0                      0      1599
```

```
6535                                     0                  0  1599
6536                                     0                  0  1599
6537                                     0                  0  1599
6538                                     0                  0  1599


      lumpedpreordersubscriptionunits  applepreorderrevenue  amazontotalunits  \
6534                                0                     0                53
6535                                0                     0                12
6536                                0                     0                25
6537                                0                     0                52
6538                                0                     0                28


      totalrevenue  barnesandnobleaccruedpreorderunits  preorderunits  \
6534         81835                                   0              0
6535         18460                                   0              0
6536         38623                                   0              0
6537         80340                                   0              0
6538         43212                                   0              0


      amazonsubscriptionrevenue  amazonlumpedpreorderrevenue  \
6534                      41860                            0
6535                      10465                            0
6536                      19435                            0
6537                      40365                            0
6538                      22425                            0


      maxbarnesandnoblelistprice  maxlistprice  barnesandnoblesoldunits  \
6534                           0          1700                        0
6535                           0          1700                        0
6536                           0          1700                        0
6537                           0          1700                        0
6538                           0          1700                        0


      maxapplelistprice  amazonaccruedpreordersubscriptionrevenue  \
6534                  0                                         0
6535                  0                                         0
6536                  0                                         0
6537                  0                                         0
6538                  0                                         0


      accruedpreordersubscriptionunits  amazonkenpread  applesoldunits  \
6534                                 0               0               0
6535                                 0               0               0
6536                                 0               0               0
6537                                 0               0               0
6538                                 0               0               0
```

```
      minamazonlistprice   minprice   bookscanreported   \
6534               1700       1599                  0
6535               1700       1599                  0
6536               1700       1599                  0
6537               1700       1599                  0
6538               1700       1599                  0


      appleaccruedpreorderunits   preorderperiod datepublished   \
6534                         0               0     2015-09-29
6535                         0               0     2015-09-29
6536                         0               0     2015-09-29
6537                         0               0     2015-09-29
6538                         0               0     2015-09-29


      lumpedpreordersoldunits   accruedpreorderrevenue   maxamazonlistprice   \
6534                        0                        0                 1700
6535                        0                        0                 1700
6536                        0                        0                 1700
6537                        0                        0                 1700
6538                        0                        0                 1700


      amazonaccruedpreorderrevenue   barnesandnoblepreorderunits   \
6534                             0                             0
6535                             0                             0
6536                             0                             0
6537                             0                             0
6538                             0                             0


      amazonaccruedpreorderunits   minbarnesandnoblelistprice   \
6534                           0                            0
6535                           0                            0
6536                           0                            0
6537                           0                            0
6538                           0                            0


      appleaccruedpreorderrevenue   applelumpedpreorderunits   \
6534                            0                          0
6535                            0                          0
6536                            0                          0
6537                            0                          0
6538                            0                          0


      applesalesrevenue   preordersoldunits   lumpedpreordersalesrevenue   \
6534                  0                   0                            0
6535                  0                   0                            0
6536                  0                   0                            0
6537                  0                   0                            0
```

```
6538                 0                   0                          0


      totalunits  maxamazonprice  amazonlumpedpreordersoldunits  \
6534          53            1599                              0
6535          12            1599                              0
6536          25            1599                              0
6537          52            1599                              0
6538          28            1599                              0


      amazonaveragerating publishertype  accruedpreorderunits  \
6534                470.0       bigfive                     0
6535                470.0       bigfive                     0
6536                470.0       bigfive                     0
6537                470.0       bigfive                     0
6538                470.0       bigfive                     0


      amazonsalesrevenue  barnesandnoblesalesrevenue  \
6534               39975                           0
6535                7995                           0
6536               19188                           0
6537               39975                           0
6538               20787                           0


                                    amazoncategories  minlistprice  \
6534  Books > Teens > Literature & Fiction > Social ...          1700
6535  Books > Teens > Literature & Fiction > Social ...          1700
6536  Books > Teens > Literature & Fiction > Social ...          1700
6537  Books > Teens > Literature & Fiction > Social ...          1700
6538  Books > Teens > Literature & Fiction > Social ...          1700


      amazonaccruedpreordersalesrevenue  amazonlumpedpreorderunits  \
6534                                  0                          0
6535                                  0                          0
6536                                  0                          0
6537                                  0                          0
6538                                  0                          0


      lumpedpreordersubscriptionrevenue  applelumpedpreorderrevenue  \
6534                                  0                           0
6535                                  0                           0
6536                                  0                           0
6537                                  0                           0
6538                                  0                           0


      minamazonprice  subscriptionunits  amazonpreorderunits  \
6534            1599                 28                    0
6535            1599                  7                    0
```

```
6536                  1599               13                      0
6537                  1599               27                      0
6538                  1599               15                      0


      publisherparent  amazonpreorderrevenue  amazonaccruedpreordersoldunits  \
6534  Simon & Schuster                      0                               0
6535  Simon & Schuster                      0                               0
6536  Simon & Schuster                      0                               0
6537  Simon & Schuster                      0                               0
6538  Simon & Schuster                      0                               0


      amazonpreordersoldunits  salesrevenue  preordersubscriptionrevenue  \
6534                        0         39975                            0
6535                        0          7995                            0
6536                        0         19188                            0
6537                        0         39975                            0
6538                        0         20787                            0


      lumpedpreorderunits  amazonaccruedpreordersubscriptionunits  \
6534                    0                                       0
6535                    0                                       0
6536                    0                                       0
6537                    0                                       0
6538                    0                                       0


              formats  applepreorderunits  subscriptionrevenue  soldunits  \
6534  Audible Audiobook                   0                41860         25
6535  Audible Audiobook                   0                10465          5
6536  Audible Audiobook                   0                19435         12
6537  Audible Audiobook                   0                40365         25
6538  Audible Audiobook                   0                22425         13


      minapplelistprice  barnesandnoblepreorderrevenue  \
6534                  0                              0
6535                  0                              0
6536                  0                              0
6537                  0                              0
6538                  0                              0


      amazonsubscriptionunits  amazonlumpedpreordersubscriptionunits  \
6534                       28                                      0
6535                        7                                      0
6536                       13                                      0
6537                       27                                      0
6538                       15                                      0


      aapreported  accruedpreordersubscriptionrevenue  foundatapple  \
```

6534	1	0	0
6535	1	0	0
6536	1	0	0
6537	1	0	0
6538	1	0	0

	barnesandnoblelumpedpreorderrevenue	barnesandnoblelumpedpreorderunits \
6534	0	0
6535	0	0
6536	0	0
6537	0	0
6538	0	0

	minbarnesandnobleprice	accruedpreordersalesrevenue \
6534	0	0
6535	0	0
6536	0	0
6537	0	0
6538	0	0

	maxbarnesandnobleprice	preorderrevenue	pubtrackreported \
6534	0	0	1
6535	0	0	1
6536	0	0	1
6537	0	0	1
6538	0	0	1

	lumpedpreorderrevenue	amazonsoldunits \
6534	0	25
6535	0	5
6536	0	12
6537	0	25
6538	0	13

	amazonpreordersubscriptionrevenue	preordersubscriptionunits \
6534	0	0
6535	0	0
6536	0	0
6537	0	0
6538	0	0

	amazonpreordersalesrevenue	maxappleprice	pagelength \
6534	0	0	395.0
6535	0	0	395.0
6536	0	0	395.0
6537	0	0	395.0
6538	0	0	395.0

	pricewassetbypublisher	barnesandnobledistributor	appledistributor
6534	NaN	NaN	NaN
6535	NaN	NaN	NaN
6536	NaN	NaN	NaN
6537	NaN	NaN	NaN
6538	NaN	NaN	NaN

```
[7]: data.isna().sum()
```

```
[7]: entity                                        0
     formattype                                    0
     fromdate                                      0
     titlecount                                    0
     titles                                        0
     todate                                        0
     author                                        0
     amazontotalreviews                           17
     minappleprice                                 0
     amazontotalrevenue                            0
     accruedpreordersoldunits                      0
     amazonlumpedpreordersubscriptionrevenue       0
     preordersalesrevenue                          0
     foundatamazon                                 0
     barnesandnobleaccruedpreorderrevenue          0
     amazonpreordersubscriptionunits               0
     amazonlumpedpreordersalesrevenue              0
     foundatbarnesandnoble                         0
     maxprice                                      0
     lumpedpreordersubscriptionunits               0
     applepreorderrevenue                          0
     amazontotalunits                              0
     totalrevenue                                  0
     barnesandnobleaccruedpreorderunits            0
     preorderunits                                 0
     amazonsubscriptionrevenue                     0
     amazonlumpedpreorderrevenue                   0
     maxbarnesandnoblelistprice                    0
     maxlistprice                                  0
     barnesandnoblesoldunits                       0
     maxapplelistprice                             0
     amazonaccruedpreordersubscriptionrevenue      0
     accruedpreordersubscriptionunits              0
     amazonkenpread                                0
     applesoldunits                                0
     minamazonlistprice                            0
     minprice                                      0
     bookscanreported                              0
```

```
appleaccruedpreorderunits                         0
preorderperiod                                     0
datepublished                                      0
lumpedpreordersoldunits                            0
accruedpreorderrevenue                             0
maxamazonlistprice                                 0
amazonaccruedpreorderrevenue                       0
barnesandnoblepreorderunits                        0
amazonaccruedpreorderunits                         0
minbarnesandnoblelistprice                         0
appleaccruedpreorderrevenue                        0
applelumpedpreorderunits                           0
applesalesrevenue                                  0
preordersoldunits                                  0
lumpedpreordersalesrevenue                         0
totalunits                                         0
maxamazonprice                                     0
amazonlumpedpreordersoldunits                      0
amazonaveragerating                               16
publishertype                                      0
accruedpreorderunits                               0
amazonsalesrevenue                                 0
barnesandnoblesalesrevenue                         0
amazoncategories                                   5
minlistprice                                       0
amazonaccruedpreordersalesrevenue                  0
amazonlumpedpreorderunits                          0
lumpedpreordersubscriptionrevenue                  0
applelumpedpreorderrevenue                         0
minamazonprice                                     0
subscriptionunits                                  0
amazonpreorderunits                                0
publisherparent                                    0
amazonpreorderrevenue                              0
amazonaccruedpreordersoldunits                     0
amazonpreordersoldunits                            0
salesrevenue                                       0
preordersubscriptionrevenue                        0
lumpedpreorderunits                                0
amazonaccruedpreordersubscriptionunits             0
formats                                            1
applepreorderunits                                 0
subscriptionrevenue                                0
soldunits                                          0
minapplelistprice                                  0
barnesandnoblepreorderrevenue                      0
amazonsubscriptionunits                            0
```

```
amazonlumpedpreordersubscriptionunits          0
aapreported                                     0
accruedpreordersubscriptionrevenue             0
foundatapple                                    0
barnesandnoblelumpedpreorderrevenue            0
barnesandnoblelumpedpreorderunits              0
minbarnesandnobleprice                          0
accruedpreordersalesrevenue                     0
maxbarnesandnobleprice                          0
preorderrevenue                                 0
pubtrackreported                                0
lumpedpreorderrevenue                           0
amazonsoldunits                                 0
amazonpreordersubscriptionrevenue              0
preordersubscriptionunits                       0
amazonpreordersalesrevenue                      0
maxappleprice                                    0
pagelength                                   3981
pricewassetbypublisher                       5008
barnesandnobledistributor                    4804
appledistributor                             5783
dtype: int64
```

Here, there are alot of NA values in 'pagelength', 'thepricewassetbypublisher', 'barnesandno-bledistributor', 'appledistributor' columns.

```
[8]: data['titles'][0]
```

[8]: "{'amazontotalreviews': 306, 'minappleprice': 0, 'amazontotalrevenue': 0,
'accruedpreordersoldunits': 0, 'amazonlumpedpreordersubscriptionrevenue': 0,
'preordersalesrevenue': 0, 'foundatamazon': 1,
'barnesandnobleaccruedpreorderrevenue': 0, 'amazonpreordersubscriptionunits': 0,
'amazonlumpedpreordersalesrevenue': 0, 'foundatbarnesandnoble': 0, 'maxprice':
0, 'lumpedpreordersubscriptionunits': 0, 'applepreorderrevenue': 0,
'amazontotalunits': 6, 'totalrevenue': 0, 'barnesandnobleaccruedpreorderunits':
0, 'preorderunits': 0, 'amazonsubscriptionrevenue': 0,
'amazonlumpedpreorderrevenue': 0, 'maxbarnesandnoblelistprice': 0,
'maxlistprice': 0, 'barnesandnoblesoldunits': 0, 'maxapplelistprice': 0,
'amazonaccruedpreordersubscriptionrevenue': 0,
'accruedpreordersubscriptionunits': 0, 'amazonkenpread': 0, 'applesoldunits': 0,
'minamazonlistprice': 0, 'minprice': 0, 'bookscanreported': 1,
'appleaccruedpreorderunits': 0, 'preorderperiod': 0, 'datepublished':
'2009-05-12', 'lumpedpreordersoldunits': 0, 'accruedpreorderrevenue': 0,
'maxamazonlistprice': 0, 'amazonaccruedpreorderrevenue': 0,
'barnesandnoblepreorderunits': 0, 'amazonaccruedpreorderunits': 0,
'minbarnesandnoblelistprice': 0, 'appleaccruedpreorderrevenue': 0,
'applelumpedpreorderunits': 0, 'applesalesrevenue': 0, 'preordersoldunits': 0,
'lumpedpreordersalesrevenue': 0, 'totalunits': 6, 'maxamazonprice': 0,
'amazonlumpedpreordersoldunits': 0, 'amazonaveragerating': 410, 'publishertype':
```

```
'bigfive', 'accruedpreorderunits': 0, 'amazonsalesrevenue': 0,
'barnesandnoblesalesrevenue': 0, 'amazoncategories': 'Books > Books on CD >
Mystery & Thrillers ; Books > Books on CD > Authors, A-Z > ( C ) > Clark, Mary
Higgins ; Books > Books on CD > Authors, A-Z ; Books > Books on CD > Authors,
A-Z > ( C ) ; Books > Books on CD ; Books > Books on CD > Literature & Fiction >
General ; Books > Books on CD > Literature & Fiction', 'minlistprice': 0,
'amazonaccruedpreordersalesrevenue': 0, 'amazonlumpedpreorderunits': 0,
'lumpedpreordersubscriptionrevenue': 0, 'applelumpedpreorderrevenue': 0,
'minamazonprice': 0, 'subscriptionunits': 0, 'amazonpreorderunits': 0,
'publisherparent': 'Simon & Schuster', 'amazonpreorderrevenue': 0,
'amazonaccruedpreordersoldunits': 0, 'amazonpreordersoldunits': 0,
'salesrevenue': 0, 'preordersubscriptionrevenue': 0, 'lumpedpreorderunits': 0,
'amazonaccruedpreordersubscriptionunits': 0, 'formats': 'Audio CD',
'applepreorderunits': 0, 'subscriptionrevenue': 0, 'soldunits': 6,
'minapplelistprice': 0, 'barnesandnoblepreorderrevenue': 0,
'amazonsubscriptionunits': 0, 'amazonlumpedpreordersubscriptionunits': 0,
'aapreported': 1, 'accruedpreordersubscriptionrevenue': 0, 'foundatapple': 0,
'barnesandnoblelumpedpreorderrevenue': 0, 'barnesandnoblelumpedpreorderunits':
0, 'minbarnesandnobleprice': 0, 'accruedpreordersalesrevenue': 0,
'maxbarnesandnobleprice': 0, 'preorderrevenue': 0, 'pubtrackreported': 1,
'lumpedpreorderrevenue': 0, 'amazonsoldunits': 6,
'amazonpreordersubscriptionrevenue': 0, 'preordersubscriptionunits': 0,
'amazonpreordersalesrevenue': 0, 'maxappleprice': 0}"
```

removing titles column as we have all the information that is included in the column

```python
[9]: data.drop('titles',axis=1,inplace=True)
```

```python
[10]: data['entity'].describe()
```

```
[10]: count                6539
      unique                  2
      top        The Walking Dead
      freq                 3499
      Name: entity, dtype: object
```

```python
[11]: data['formattype'].unique()
```

```
[11]: array(['print', 'ebook', 'audio'], dtype=object)
```

```python
[13]: # count plot of Alone in the dark and the walking dead.
      sns.countplot('entity',data=data)
```

```
[13]: <matplotlib.axes._subplots.AxesSubplot at 0x1db4f8390b8>
```
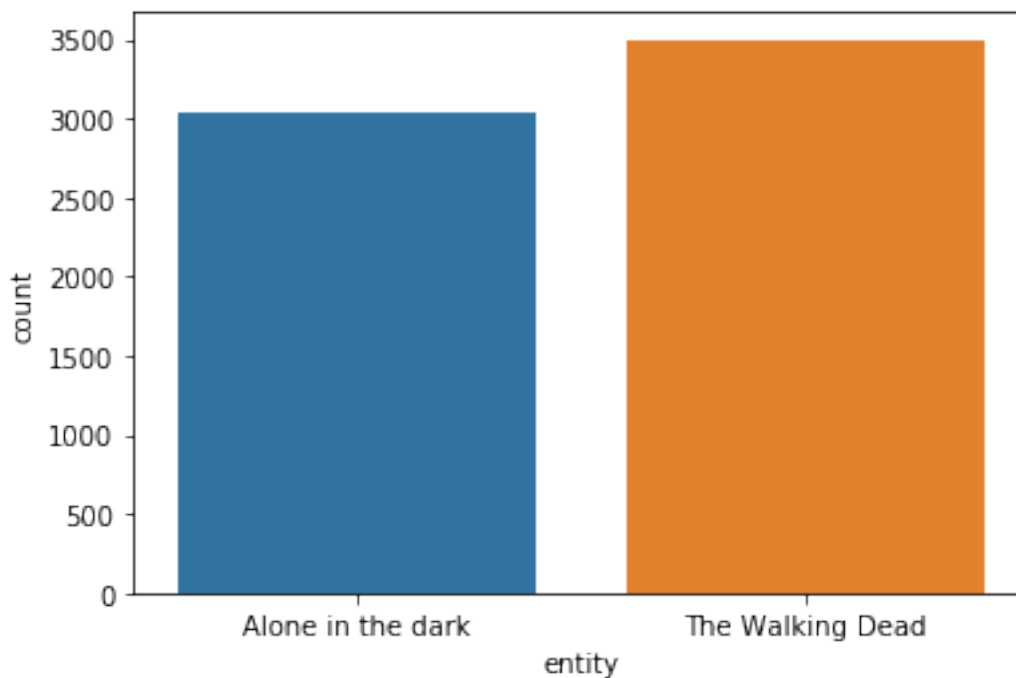
```python
#count plot of format type.
sns.countplot('formattype',data=data)
```

`<matplotlib.axes._subplots.AxesSubplot at 0x1db4fb96470>`
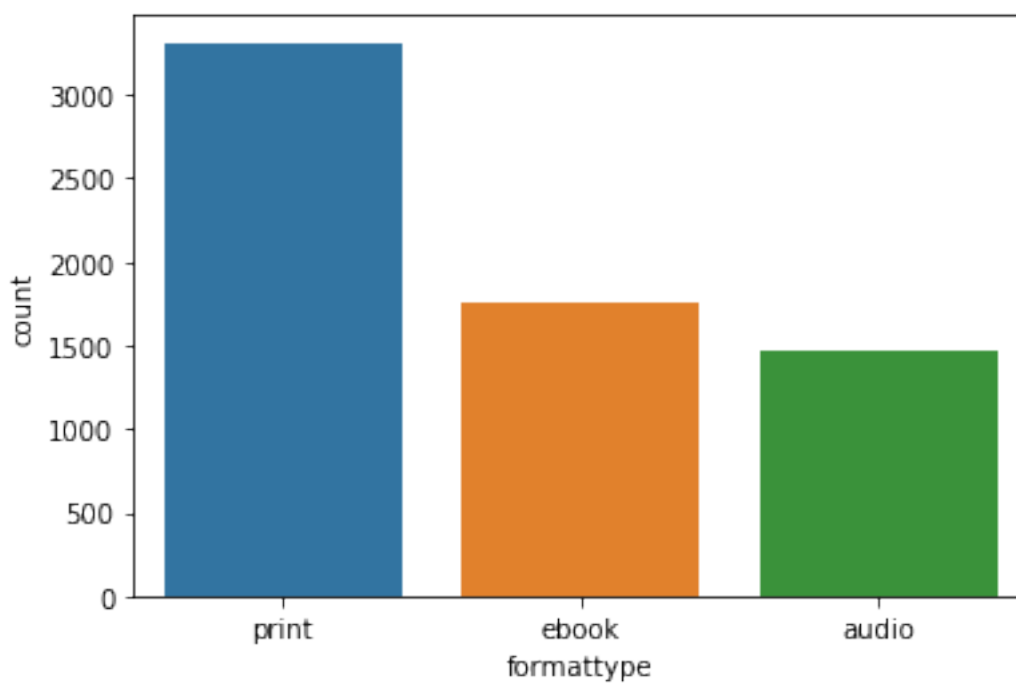
```
[15]: df_entity = data.groupby('entity')
```

```
[16]: The_walking_dead = df_entity.get_group('The Walking Dead')
```

```
[17]: Alone_in_the_dark = df_entity.get_group('Alone in the dark')
```

```
[18]: # checking total count comparision of Alone in The dark by Format type.
      sns.countplot('formattype',data= Alone_in_the_dark)
```

[18]: <matplotlib.axes._subplots.AxesSubplot at 0x1db500be438>
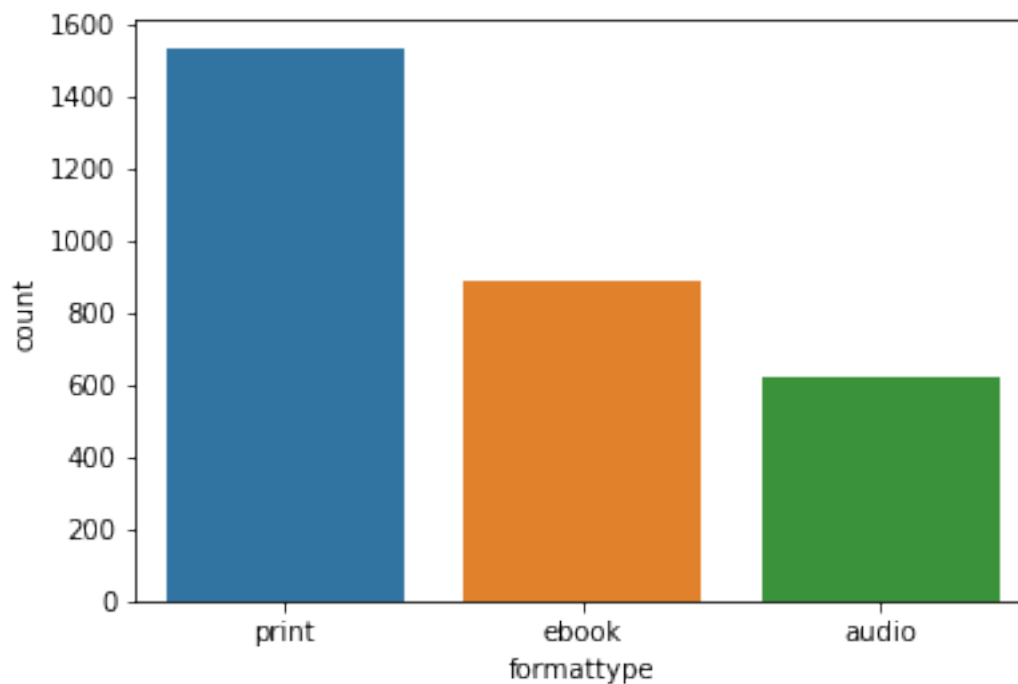

```
[19]: # checking total count comparision of The Walking Dead by Format type.
      sns.countplot('formattype',data= The_walking_dead)
```

[19]: <matplotlib.axes._subplots.AxesSubplot at 0x1db500a7b70>
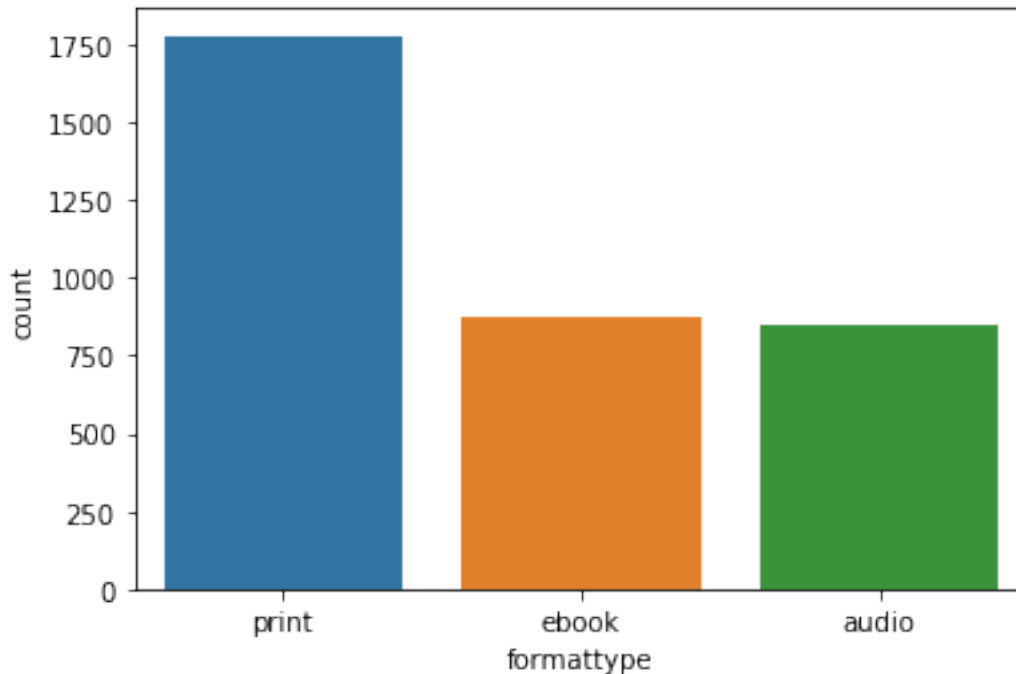
```
[20]: Print = data[data['formattype']=='print']
```

```
[21]: Print['totalrevenue'].sum()
```

```
[21]: 72444324
```

```
[22]: Ebook = data[data['formattype']=='ebook']
```

```
[23]: Ebook['totalrevenue'].sum()
```

```
[23]: 21636050
```

```
[24]: Audio = data[data['formattype']=='audio']
```

```
[25]: Audio['totalrevenue'].sum()
```

```
[25]: 30469078
```

```
[26]: Revenue = pd.DataFrame(data= [72444324,␣
      ↪21636050,30469078],index=['Print','Ebook','Audio'],columns=['Total_Revenue'])
```

```
[27]: Revenue = Revenue.reset_index()
```

```
[28]: Revenue.columns = ['format_type', 'Total_Revenue']
```

```
[29]: # checking Toatl revenue comparision by format type.
      sns.barplot('format_type','Total_Revenue',data= Revenue)
```

```
[29]: <matplotlib.axes._subplots.AxesSubplot at 0x1db500dc898>
```
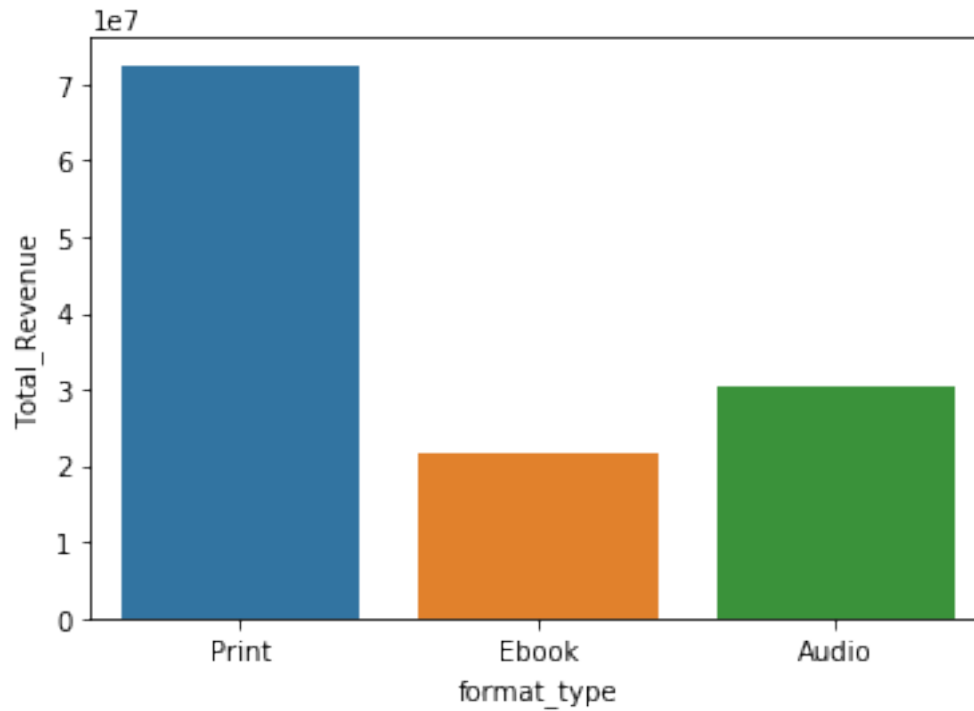
```
[30]: data['amazonsalesrevenue'].sum()
```

```
[30]: 102739496
```

```
[31]: data['applesalesrevenue'].sum()
```

```
[31]: 6341652
```

```
[32]: data['barnesandnoblesalesrevenue'].sum()
```

```
[32]: 15984
```

```
[33]: sales_revenue =  pd.DataFrame(data= [
      102739496,␣
       ↪6341652,15984],index=['Amazon','Apple','BarnesNoble'],columns=['Sales_revenue'])
```

```
[34]: sales_revenue = sales_revenue.reset_index()
      sales_revenue.columns = ['Sold_by','Sales_revenue']
```

```
[35]: # checking sales revenue comparision of amazon,apple & BarnesandNoble
      sns.barplot('Sold_by','Sales_revenue',data = sales_revenue)
```

```
[35]: <matplotlib.axes._subplots.AxesSubplot at 0x1db5061ed68>
```
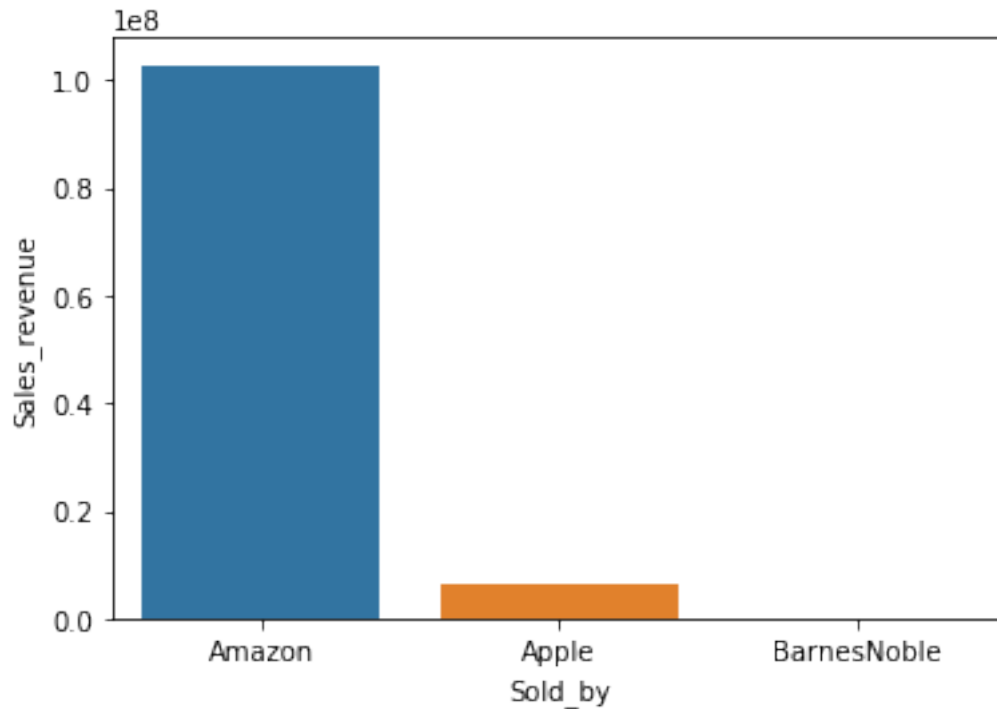
```
[36]: sold_units =␣
      →data[data['formattype']=='print'][['amazonsoldunits','barnesandnoblesoldunits','applesoldun
```

```
[37]: sold_units.sum() # we can see that there has been no units sold of print format␣
      →type by apple and barnesandnoble.
```

```
[37]: amazonsoldunits              88360
      barnesandnoblesoldunits          0
      applesoldunits                   0
      dtype: int64
```

```
[38]: units_by_print =  pd.DataFrame(data= [88360,0,0],index=['amazonsoldunits␣
      →','barnesandnoblesoldunits','applesoldunits '],columns=['Total_units'])
```

```
[39]: #units for print format
      plt.figure(dpi=100)
      explode_list= [0.0,0.1,0.2]
      color_list = ['lightblue','brown','green']
      units_by_print['Total_units'].plot(kind='pie',
                          figsize=(10,6),
                          autopct='(%1.2f%%)',
                           startangle= 45,
                        colors = color_list,
                        pctdistance = 1.37,
                        explode = explode_list)
      plt.title('units for print Distribution')
```

```python
plt.legend(units_by_print.index,loc = 'upper left')
```

[39]: <matplotlib.legend.Legend at 0x1db506ba710>

units for print Distribution
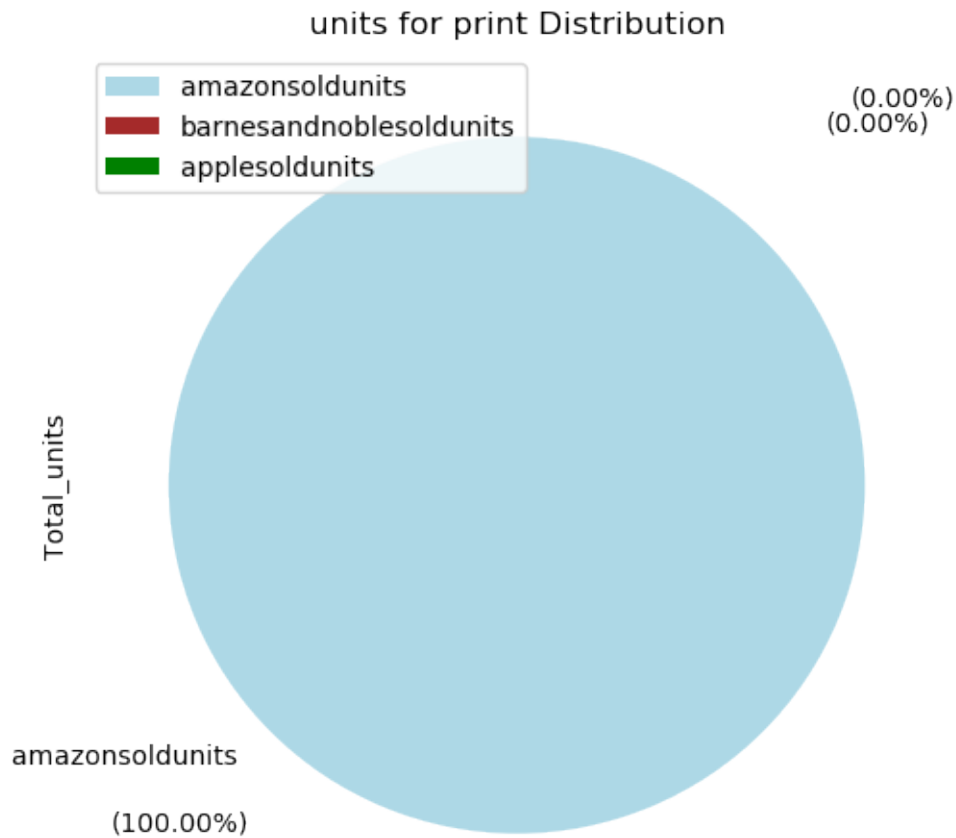
[40]:
```python
sold_units =␣
 ↪data[data['formattype']=='ebook'][['amazonsoldunits','barnesandnoblesoldunits','applesoldun
```

[41]:
```python
sold_units.sum() # we can see that there were very few units sold of ebook␣
 ↪format type by barnesandnoble compared to rest 2.
```

[41]:
```
amazonsoldunits            17806
barnesandnoblesoldunits       16
applesoldunits              6348
dtype: int64
```

[42]:
```python
units_by_ebook =  pd.DataFrame(data= [17806,16,6348],index=['amazonsoldunits␣
 ↪','barnesandnoblesoldunits','applesoldunits '],columns=['Total_units'])
```

[43]:
```python
#units for print format
plt.figure(dpi=100)
explode_list= [0.0,0.05,0.1]
```

```
color_list = ['lightblue','brown','green']
units_by_ebook['Total_units'].plot(kind='pie',
                        figsize=(10,6),
                        autopct='(%1.1f%%)',
                         startangle= 20,
                    colors = color_list,
                    pctdistance = 1.37,
                    explode = explode_list)
plt.title('units for ebook Distribution')
plt.legend(units_by_ebook.index,loc = 'upper left')
```

[43]: <matplotlib.legend.Legend at 0x1db509c7e48>

units for ebook Distribution
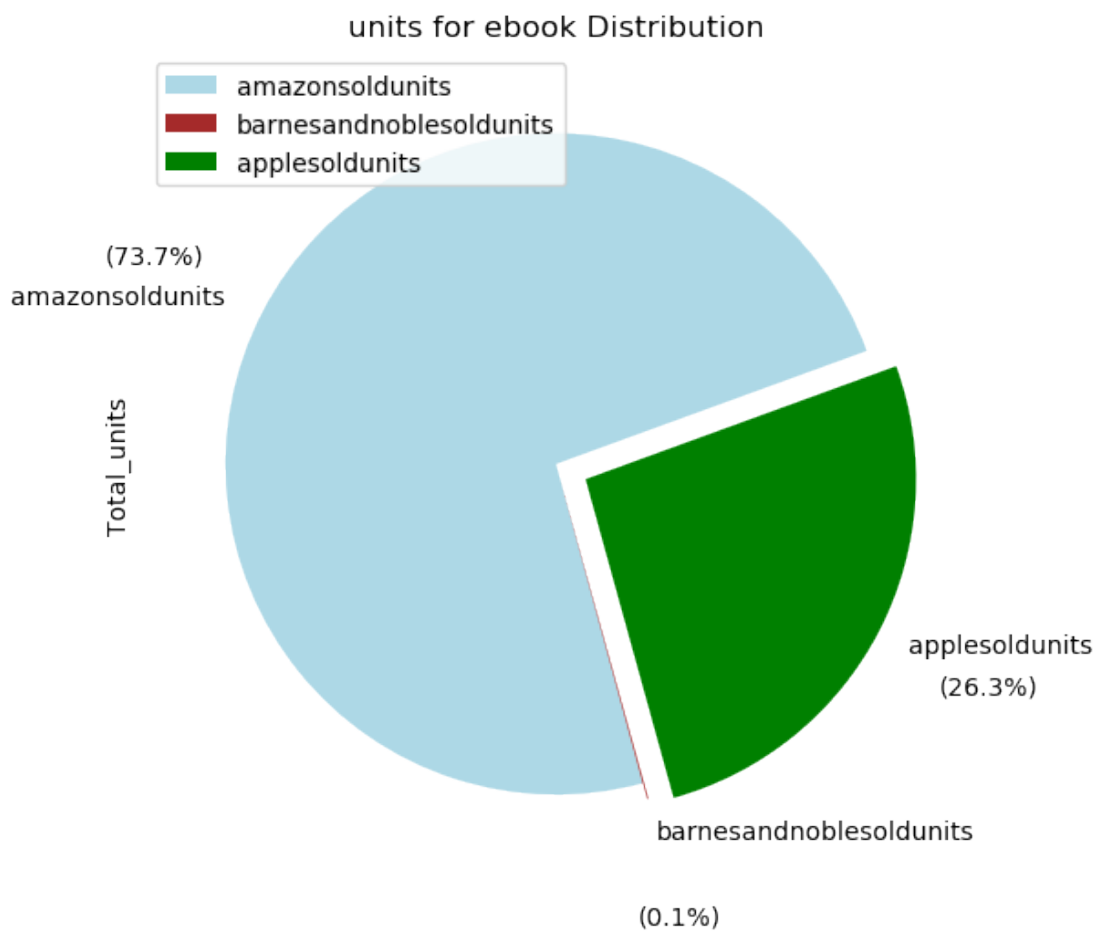
[44]:
```
sold_units =␣
 ↪data[data['formattype']=='audio'][['amazonsoldunits','barnesandnoblesoldunits','applesoldun
```

[45]:
```
sold_units.sum() # we can see that there has been no units sold of audio format␣
 ↪type by apple and barnesandnoble.
```

```
[45]:  amazonsoldunits              9352
       barnesandnoblesoldunits         0
       applesoldunits                  0
       dtype: int64
```

```
[46]:  units_by_audio =  pd.DataFrame(data= [9352,0,0],index=['amazonsoldunits␣
        ↪','barnesandnoblesoldunits','applesoldunits '],columns=['Total_units'])
```

```
[47]:  #units for print format
       plt.figure(dpi=100)
       explode_list= [0.0,0.1,0.2]
       color_list = ['lightblue','brown','green']
       units_by_audio['Total_units'].plot(kind='pie',
                             figsize=(10,6),
                             autopct='(%1.1f%%)',
                              startangle= 50,
                         colors = color_list,
                         pctdistance = 1.37,
                         explode = explode_list)
       plt.title('units for audio Distribution')
       plt.legend(units_by_audio.index,loc = 'upper left')
```

```
[47]:  <matplotlib.legend.Legend at 0x1db50787e80>
```

units for audio Distribution
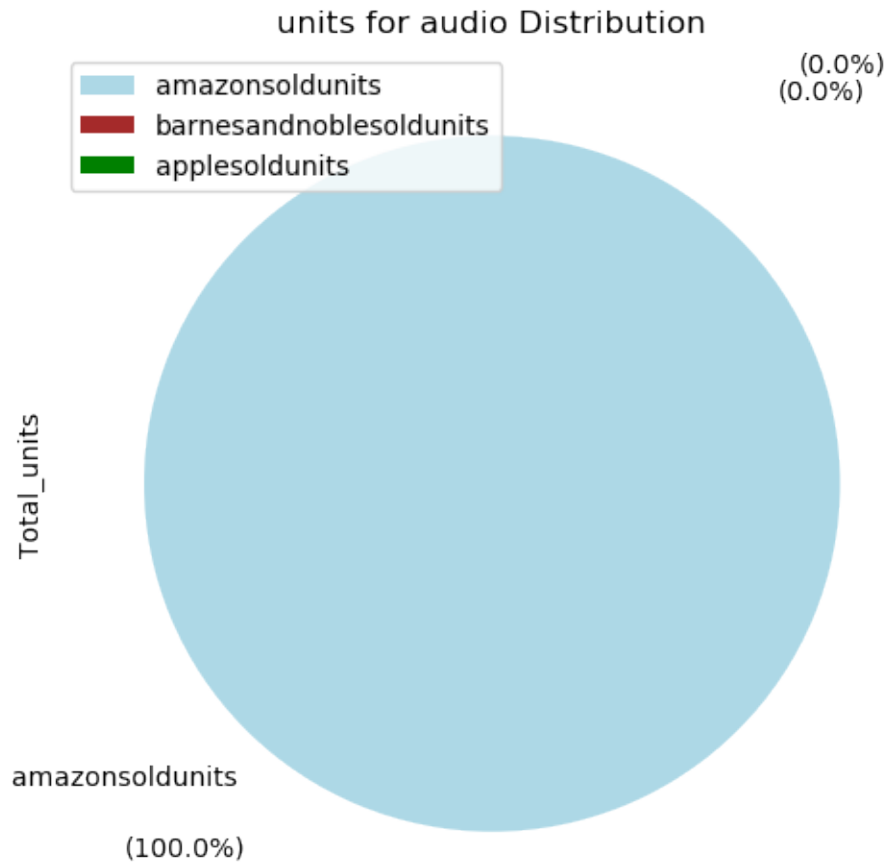

```
[48]: Total_units_by_formattype = data[['formattype','totalunits']]
```

```
[49]: total_units = Total_units_by_formattype.groupby('formattype')
```

```
[50]: print_units = total_units.get_group('print').sum()
      print_units
```

```
[50]: formattype    printprintprintprintprintprintprintprintprintp...
      totalunits                                               88360
      dtype: object
```

```
[51]: ebook_units = total_units.get_group('ebook').sum()
      ebook_units
```

```
[51]: formattype    ebookebookebookebookebookebookebookebookebooke...
      totalunits                                              24170
      dtype: object
```

```
[52]: audio_units = total_units.get_group('audio').sum()
      audio_units
```

[52]: formattype   audioaudioaudioaudioaudioaudioaudioaudioaudioa...
      totalunits   19688
      dtype: object

[53]:
```
units_by_formattype =  pd.DataFrame(data= [88360,
↪24170,19688],index=['Print','Ebook','Audio'],columns=['Total_units'])
```

[54]:
```
#units by format type.
plt.figure(dpi=100)
explode_list= [0.0,0.1,0.2]
color_list = ['lightblue','brown','green']
units_by_formattype['Total_units'].plot(kind='pie',
                       figsize=(10,6),
                       autopct='(%1.2f%%)',
                        startangle= 40,
                    colors = color_list,
                    pctdistance = 1.37,
                    explode = explode_list)
plt.title('units by formattype Distribution')
plt.legend(units_by_formattype.index,loc = 'upper left')
```

[54]: <matplotlib.legend.Legend at 0x1db500ff6d8>
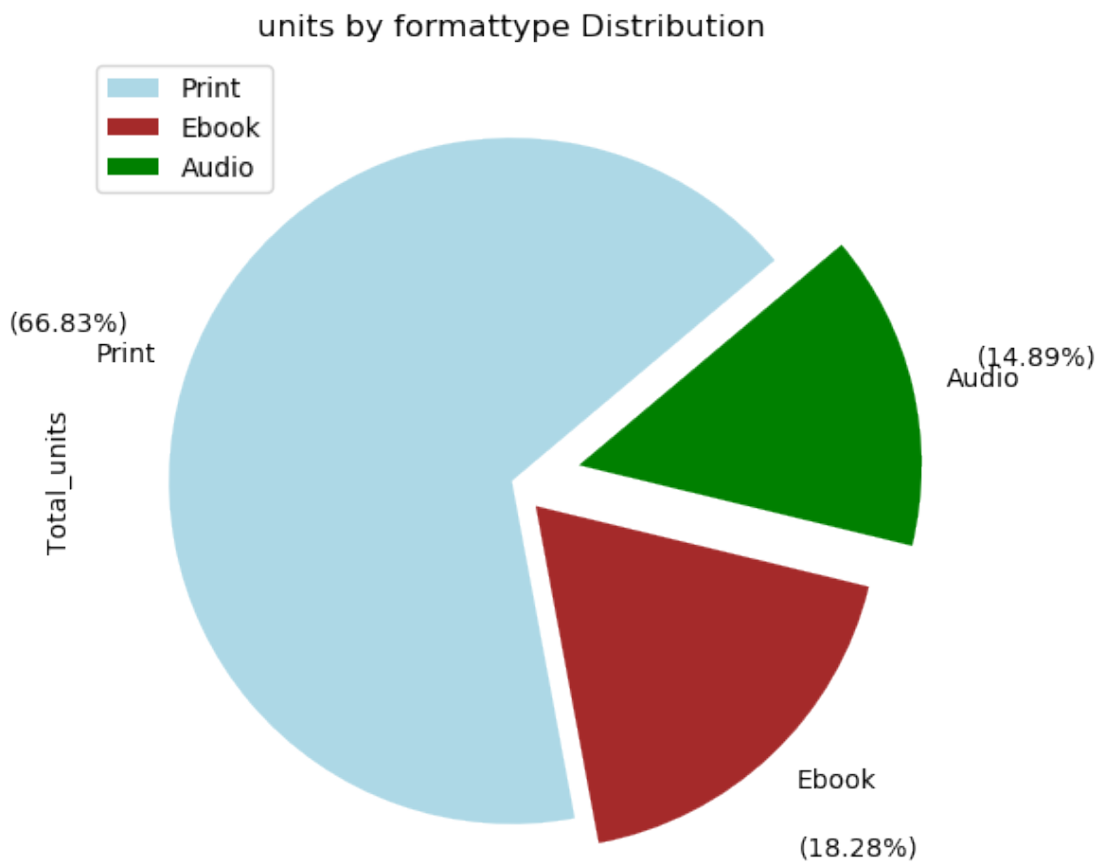
Most number of units are sold by print format type.

```
[55]: subscriptions = data[['subscriptionrevenue','totalrevenue']]
```

```
[56]: subscriptions['subscriptionrevenue'].sum()
```

```
[56]: 15452320
```

```
[57]: subscriptions['totalrevenue'].sum()
```

```
[57]: 124549452
```

```
[58]: revenue_sub =  pd.DataFrame(data= [15452320,␣
       ↪124549452],index=['subscriptionrevenue','totalrevenue'],columns=['revenue'])
```

```
[59]: revenue_sub  = revenue_sub.reset_index()
      revenue_sub.columns = ['comp','revenue']
```

```
[60]: #comparision between revenue by subscriptions and total_revenue.
      sns.barplot('comp','revenue',data = revenue_sub)
```

```
[60]: <matplotlib.axes._subplots.AxesSubplot at 0x1db5098ceb8>
```
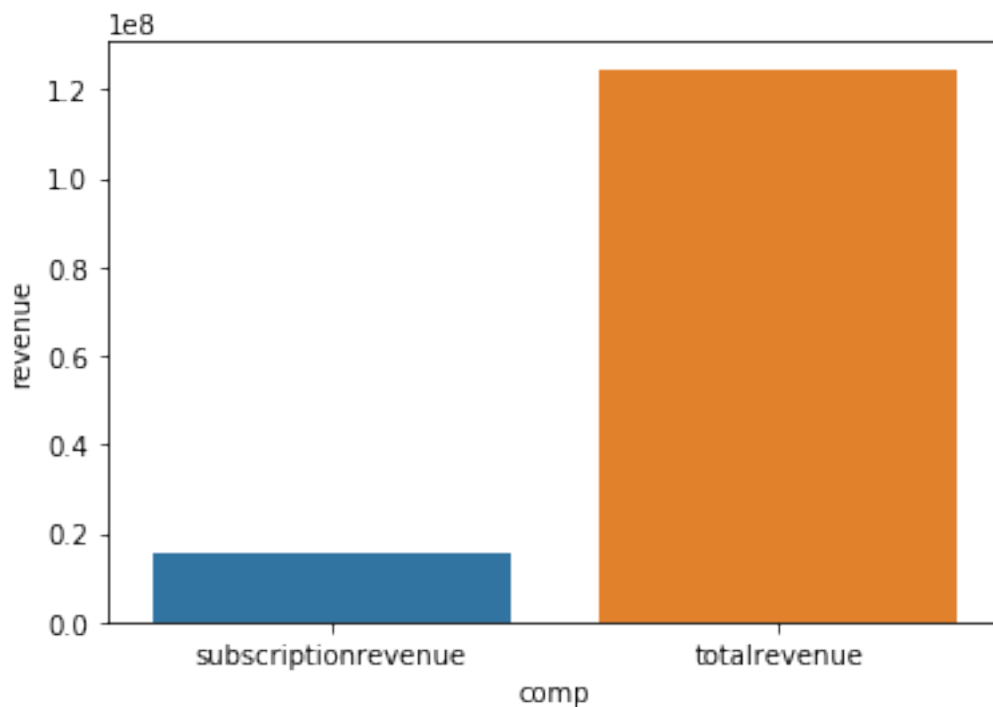


```
[61]: units_sold_resp_entity = data[['entity','soldunits']]
```

```python
[62]: units_sold_resp_entity[units_sold_resp_entity['entity']=='Alone in the dark'].
      ↪sum()
```

```
[62]: entity        Alone in the darkAlone in the darkAlone in the...
      soldunits                                              11549
      dtype: object
```

```python
[63]: units_sold_resp_entity[units_sold_resp_entity['entity']=='The Walking Dead'].
      ↪sum()
```

```
[63]: entity        The Walking DeadThe Walking DeadThe Walking De...
      soldunits                                              110333
      dtype: object
```

```python
[64]: units_sold_resp_Entity =  pd.DataFrame(data= [11549, 110333],index=['Alone in␣
      ↪the dark','The Walking Dead'],columns=['sold units'])
```

```python
[65]: units_sold_resp_Entity  = units_sold_resp_Entity.reset_index()
      units_sold_resp_Entity.columns = ['Entity','sold units']
```

```python
[66]: # Number of units sold by each entity.
      sns.barplot('Entity','sold units',data = units_sold_resp_Entity)
```

```
[66]: <matplotlib.axes._subplots.AxesSubplot at 0x1db50a515c0>
```
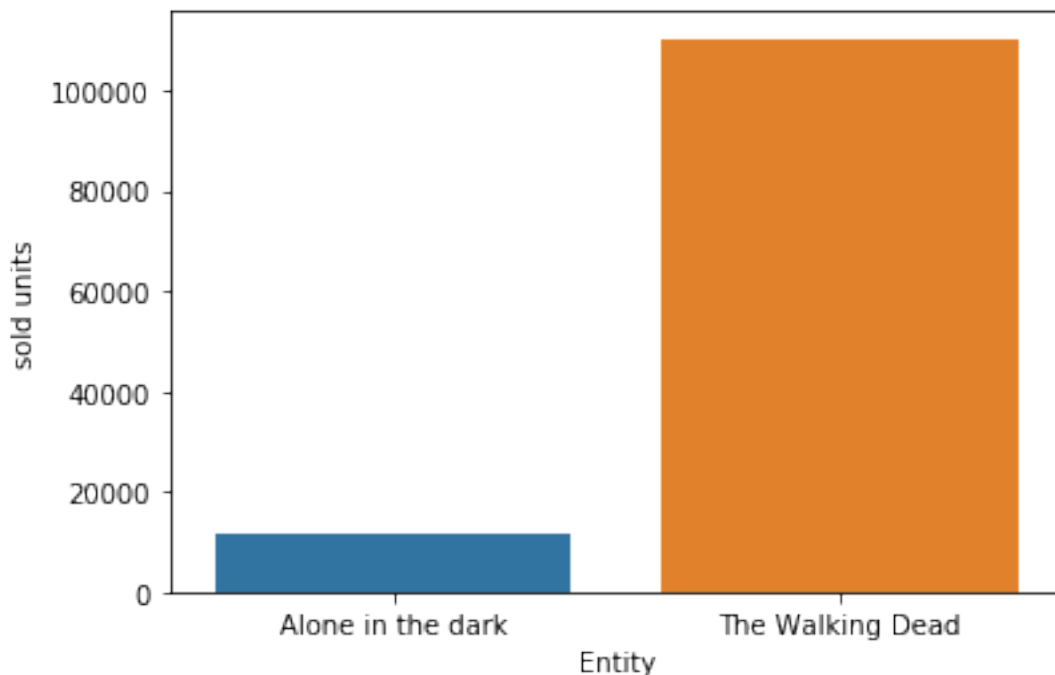


```python
[67]: revenue_resp_entity = data[['entity','totalrevenue']]
```

```python
[68]: revenue_resp_entity[revenue_resp_entity['entity']=='Alone in the dark'].sum()
```

```
[68]: entity            Alone in the darkAlone in the darkAlone in the...
      totalrevenue                                             10302489
      dtype: object
```

```
[69]: revenue_resp_entity[revenue_resp_entity['entity']=='The Walking Dead'].sum()
```

```
[69]: entity            The Walking DeadThe Walking DeadThe Walking De...
      totalrevenue                                            114246963
      dtype: object
```

```
[70]: revenue_resp_Entity =  pd.DataFrame(data= [10302489, 114246963],index=['Alone␣
      ↪in the dark','The Walking Dead'],columns=['total revenue'])
```

```
[71]: plt.figure(dpi=100)
      explode_list= [0.0,0.2]
      color_list = ['lightblue','brown']
      revenue_resp_Entity['total revenue'].plot(kind='pie',
                              figsize=(10,6),
                              autopct='(%1.2f%%)',
                               startangle= 45,
                         colors = color_list,
                         pctdistance = 1.28,
                         explode = explode_list)
      plt.title('Entity revenue Distribution')
      plt.legend(revenue_resp_Entity.index,loc = 'upper left')
```

```
[71]: <matplotlib.legend.Legend at 0x1db50aa7f60>
```

Entity revenue Distribution
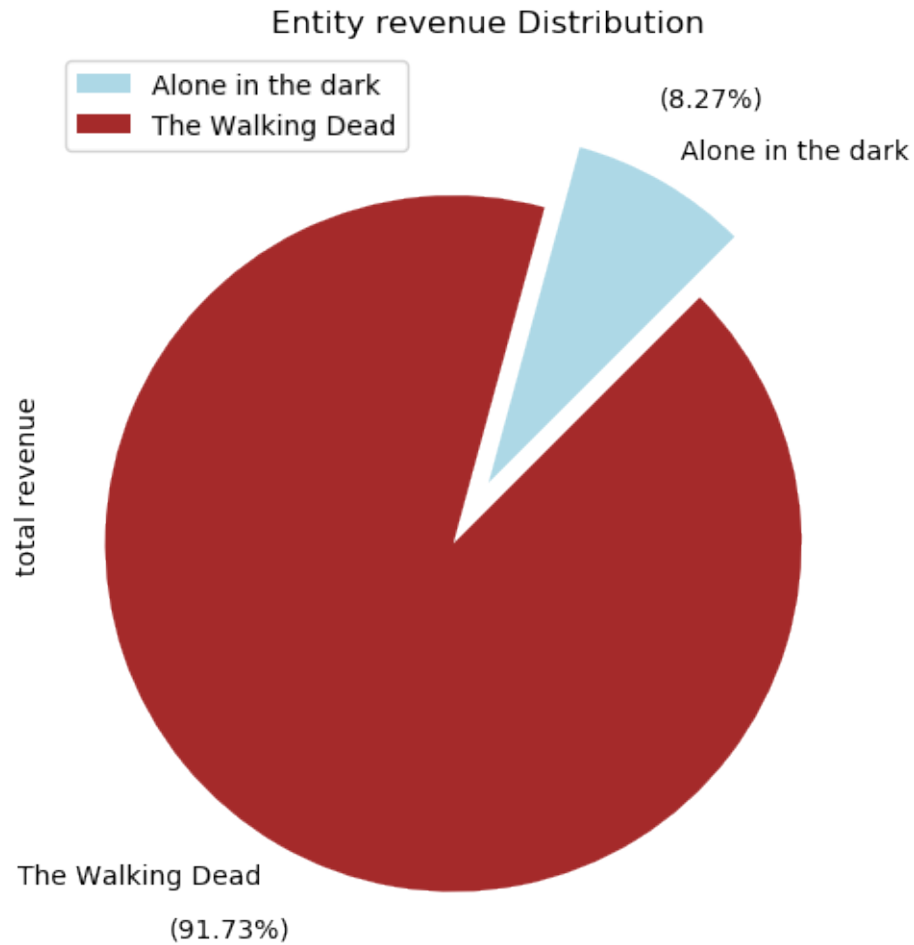
We can see that most of the revenue came from The walking Dead.

```
[72]: Amazon_revenue =␣
      ↪data[['amazontotalrevenue','amazonlumpedpreordersubscriptionrevenue','amazonlumpedpreorders
            ␣
      ↪'amazonsubscriptionrevenue','amazonlumpedpreorderrevenue','amazonaccruedpreordersubscriptio
            ␣
      ↪,'amazonaccruedpreorderrevenue','amazonsalesrevenue','amazonaccruedpreordersalesrevenue',
            ␣
      ↪'amazonpreordersubscriptionrevenue','amazonpreordersalesrevenue']]
```

```
[73]: Amazon_revenue.sum()
```

```
[73]: amazontotalrevenue                      118191816
      amazonlumpedpreordersubscriptionrevenue         0
      amazonlumpedpreordersalesrevenue          1016836
      amazonsubscriptionrevenue                15452320
      amazonlumpedpreorderrevenue               1016836
```

```
amazonaccruedpreordersubscriptionrevenue                    0
amazonaccruedpreorderrevenue                          1193792
amazonsalesrevenue                                  102739496
amazonaccruedpreordersalesrevenue                     1193792
amazonpreordersubscriptionrevenue                           0
amazonpreordersalesrevenue                            1016836
dtype: int64
```

[74]:
```python
amazon_revenue =  pd.DataFrame(data=␣
 ↪[118191816,0,1016836,15452320,1016836,0,1193792,102739496,1193792,0,1016836],index=['amazon
                         ␣
 ↪'amazonsubscriptionrevenue','amazonlumpedpreorderrevenue','amazonaccruedpreordersubscriptio
                         ␣
 ↪,'amazonaccruedpreorderrevenue','amazonsalesrevenue','amazonaccruedpreordersalesrevenue',
                         ␣
 ↪'amazonpreordersubscriptionrevenue','amazonpreordersalesrevenue'],columns=['total␣
 ↪revenue'])
```

[75]:
```python
amazon_revenue  = amazon_revenue.reset_index()
amazon_revenue.columns = ['Resources','total_revenue']
```

[76]:
```python
# total revenue generated by various types of resources at amazon.
chart = sns.barplot('Resources','total_revenue',data = amazon_revenue)
chart.set_xticklabels(chart.get_xticklabels(), rotation=45,␣
 ↪horizontalalignment='right')
```

[76]:
```
[Text(0, 0, 'amazontotalrevenue'),
 Text(0, 0, 'amazonlumpedpreordersubscriptionrevenue'),
 Text(0, 0, 'amazonlumpedpreordersalesrevenue'),
 Text(0, 0, 'amazonsubscriptionrevenue'),
 Text(0, 0, 'amazonlumpedpreorderrevenue'),
 Text(0, 0, 'amazonaccruedpreordersubscriptionrevenue'),
 Text(0, 0, 'amazonaccruedpreorderrevenue'),
 Text(0, 0, 'amazonsalesrevenue'),
 Text(0, 0, 'amazonaccruedpreordersalesrevenue'),
 Text(0, 0, 'amazonpreordersubscriptionrevenue'),
 Text(0, 0, 'amazonpreordersalesrevenue')]
```
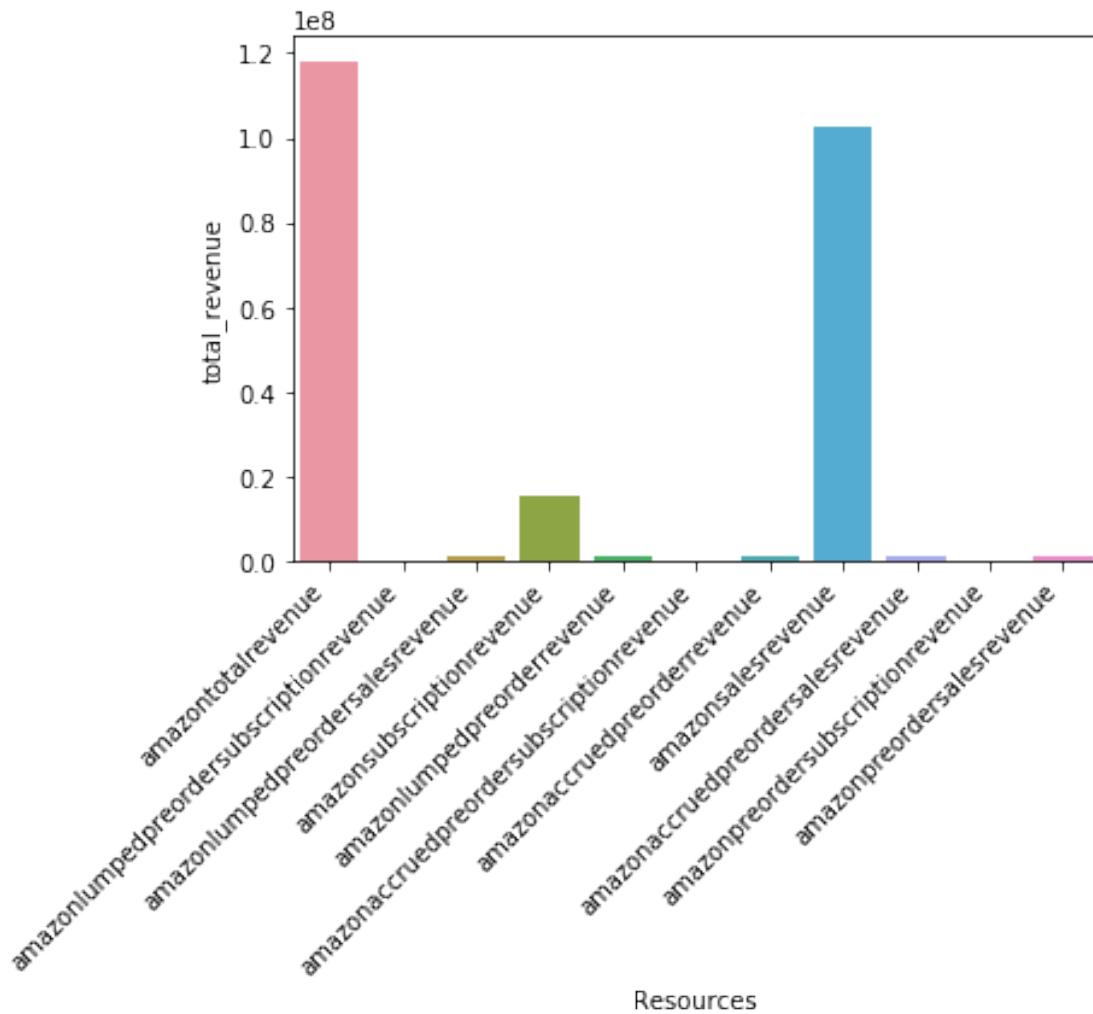
Most of the amazon revenue came from amazonsubscriptionrevenue and amazonsalesrevenue.

```
[77]: Apple_revenue = data[['applepreorderrevenue','appleaccruedpreorderrevenue',
      ↳'applesalesrevenue','applelumpedpreorderrevenue']]
```

```
[78]: Apple_revenue.sum()
```

```
[78]: applepreorderrevenue              0
      appleaccruedpreorderrevenue       0
      applesalesrevenue           6341652
      applelumpedpreorderrevenue        0
      dtype: int64
```

```
[79]: apple_revenue = pd.DataFrame(data= [0,0,6341652,0],
      ↳index=['applepreorderrevenue','appleaccruedpreorderrevenue',
```

```
                                      ⊔
    ↪'applesalesrevenue','applelumpedpreorderrevenue'],columns=['total revenue'])
```

```
[80]: apple_revenue  = apple_revenue.reset_index()
      apple_revenue.columns = ['Resources','total_revenue']
```

```
[81]: # total revenue generated by various types of resources at apple.
      chart = sns.barplot('Resources','total_revenue',data = apple_revenue)
      chart.set_xticklabels(chart.get_xticklabels(), rotation=45,⊔
      ↪horizontalalignment='right')
```

```
[81]: [Text(0, 0, 'applepreorderrevenue'),
       Text(0, 0, 'appleaccruedpreorderrevenue'),
       Text(0, 0, 'applesalesrevenue'),
       Text(0, 0, 'applelumpedpreorderrevenue')]
```
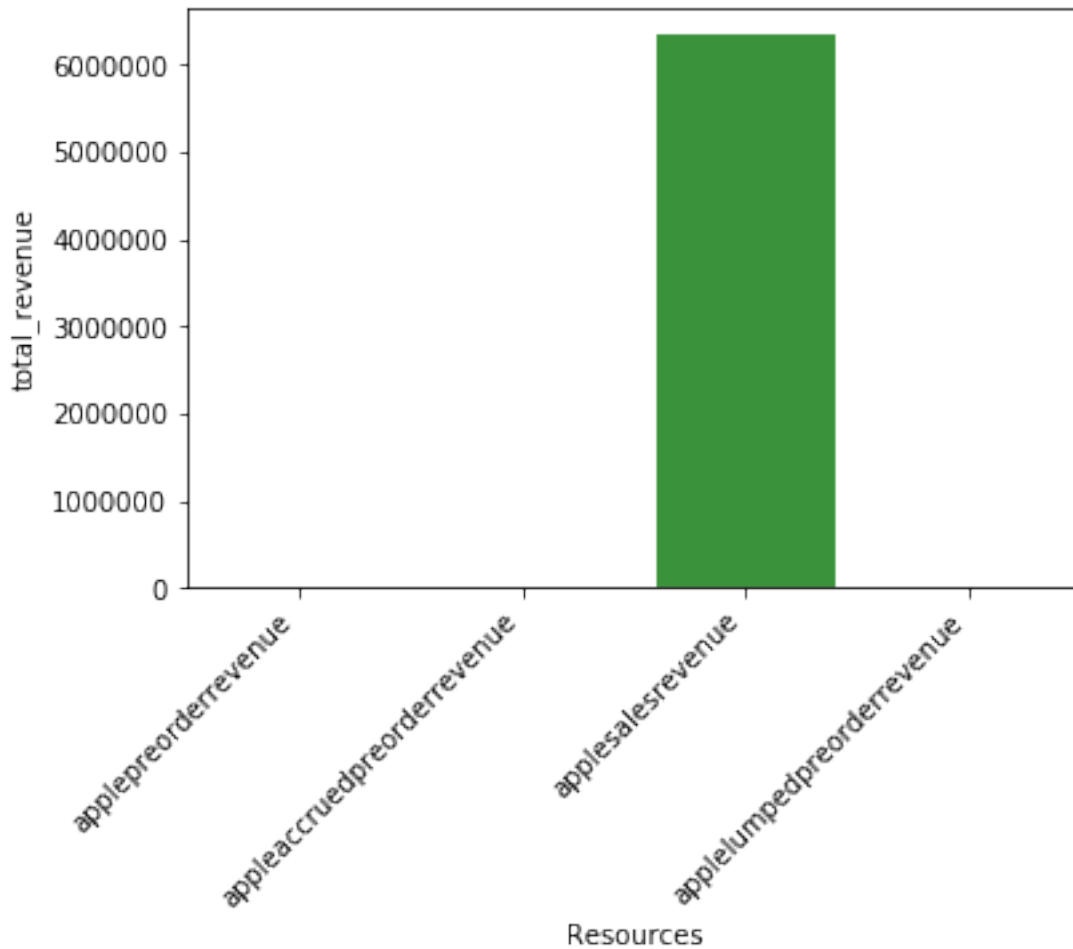

All the apple revenue is from applesalesrevenue.

```
[82]: BarnesNoble_revenue =␣
      ↪data[['barnesandnobleaccruedpreorderrevenue','barnesandnoblesalesrevenue','barnesandnoblelu
```

```
[83]: BarnesNoble_revenue.sum()
```

```
[83]: barnesandnobleaccruedpreorderrevenue             0
      barnesandnoblesalesrevenue                   15984
      barnesandnoblelumpedpreorderrevenue              0
      dtype: int64
```

```
[84]: barnesnoble_revenue = pd.DataFrame(data= [0,15984,0],
                                  ␣
      ↪index=['barnesandnobleaccruedpreorderrevenue','barnesandnoblesalesrevenue',
                              ␣
      ↪'barnesandnoblelumpedpreorderrevenue'],columns=['total revenue'])
```

```
[85]: barnesnoble_revenue  = barnesnoble_revenue.reset_index()
      barnesnoble_revenue.columns = ['Resources','total_revenue']
```

```
[86]: # total revenue generated by various types of resources at barnesandnoble.
      chart = sns.barplot('Resources','total_revenue',data = barnesnoble_revenue)
      chart.set_xticklabels(chart.get_xticklabels(), rotation=45,␣
      ↪horizontalalignment='right')
```

```
[86]: [Text(0, 0, 'barnesandnobleaccruedpreorderrevenue'),
       Text(0, 0, 'barnesandnoblesalesrevenue'),
       Text(0, 0, 'barnesandnoblelumpedpreorderrevenue')]
```
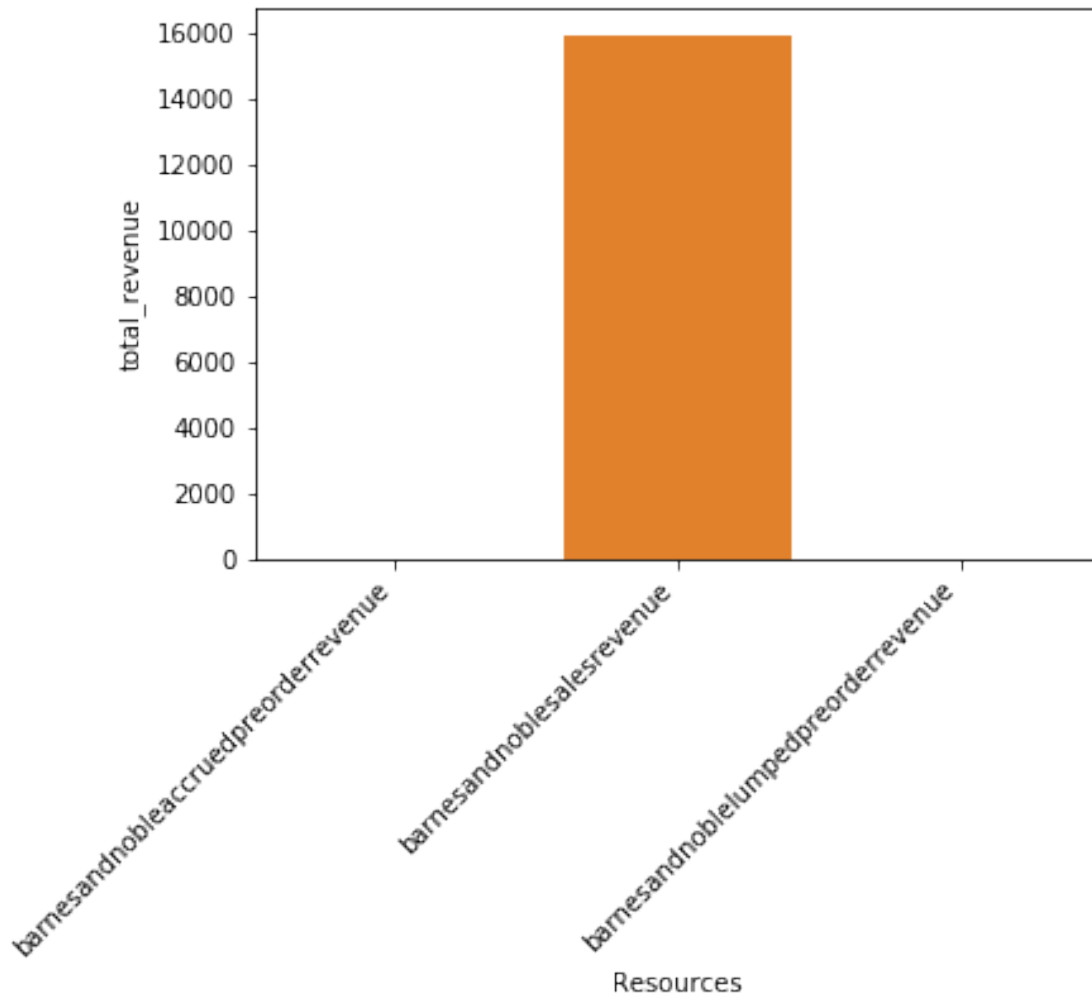
All the BarnesandNoble revenue is generated by barnesandnoblesalesrevenue.

```
[87]: Amazon_units = data[['amazontotalunits','amazonlumpedpreordersubscriptionunits',
         ⊔
      ↪'amazonsubscriptionunits','amazonlumpedpreorderunits','amazonaccruedpreordersubscriptionunit
         ⊔
      ↪,'amazonaccruedpreorderunits','amazonaccruedpreordersoldunits','amazonpreordersoldunits','a
```

```
[88]: Amazon_units.sum()
```

```
[88]: amazontotalunits                            125854
      amazonlumpedpreordersubscriptionunits            0
      amazonsubscriptionunits                      10336
      amazonlumpedpreorderunits                     1374
      amazonaccruedpreordersubscriptionunits           0
      amazonaccruedpreorderunits                    1538
      amazonaccruedpreordersoldunits                1538
      amazonpreordersoldunits                       1374
```

44

```
amazonsoldunits                                       115518
dtype: int64
```

[89]:
```python
amazon_units =  pd.DataFrame(data=␣
 ↪[125854,0,10336,1374,0,1538,1538,1374,115518],index=['amazontotalunits','amazonlumpedpreord
                   ␣
 ↪'amazonsubscriptionunits','amazonlumpedpreorderunits','amazonaccruedpreordersubscriptionuni
                  ␣
 ↪,'amazonaccruedpreorderunits','amazonaccruedpreordersoldunits','amazonpreordersoldunits','a
 ↪of units sold'])
```

[90]:
```python
amazon_units  = amazon_units.reset_index()
amazon_units.columns = ['Resources','no of units sold']
```

[91]:
```python
# total units sold by various types of resources at amazon.
chart = sns.barplot('Resources','no of units sold',data = amazon_units)
chart.set_xticklabels(chart.get_xticklabels(), rotation=45,␣
 ↪horizontalalignment='right')
```

[91]:
```
[Text(0, 0, 'amazontotalunits'),
 Text(0, 0, 'amazonlumpedpreordersubscriptionunits'),
 Text(0, 0, 'amazonsubscriptionunits'),
 Text(0, 0, 'amazonlumpedpreorderunits'),
 Text(0, 0, 'amazonaccruedpreordersubscriptionunits'),
 Text(0, 0, 'amazonaccruedpreorderunits'),
 Text(0, 0, 'amazonaccruedpreordersoldunits'),
 Text(0, 0, 'amazonpreordersoldunits'),
 Text(0, 0, 'amazonsoldunits')]
```
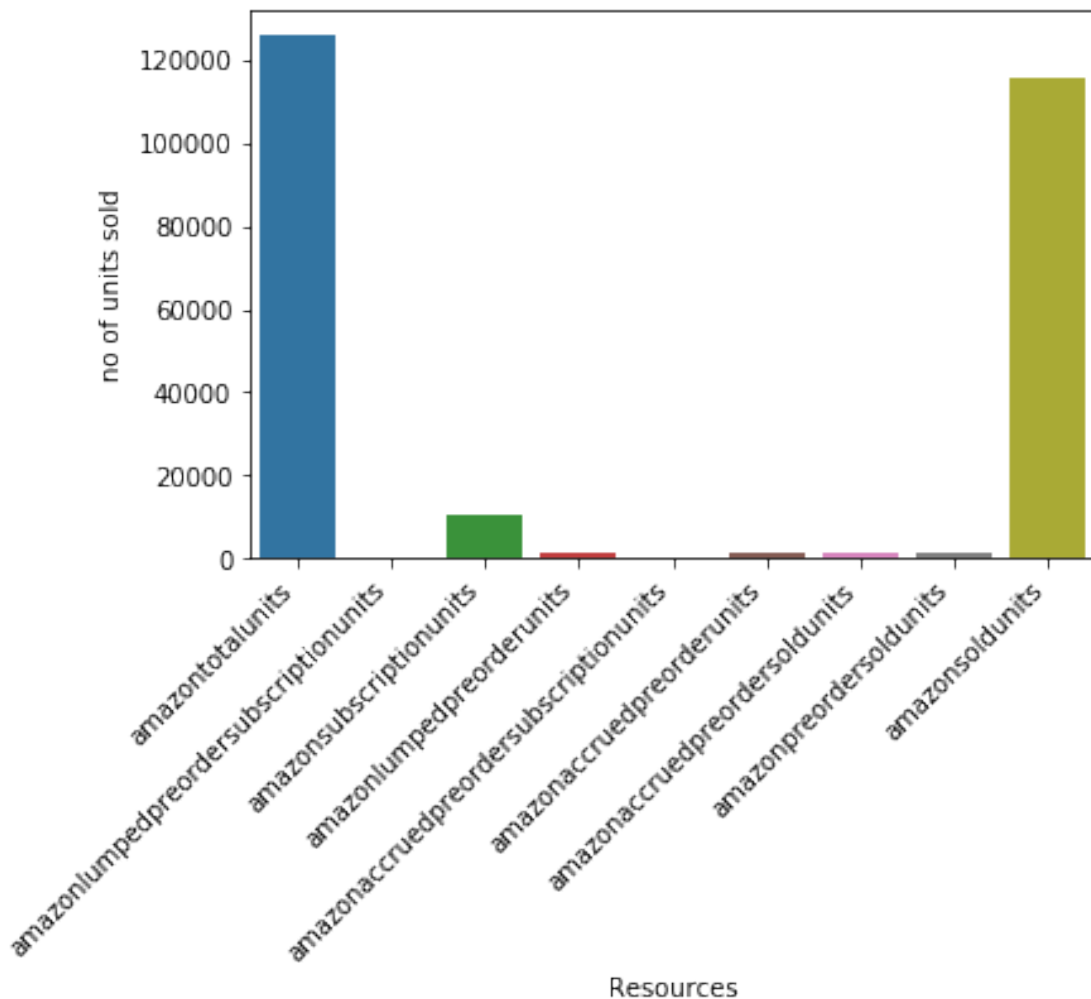
Most of the units are sold online compared to subscribtion units at amazon

```
[92]: Apple_units =␣
      ↪data[['applesoldunits','appleaccruedpreorderunits','applelumpedpreorderunits']]
```

```
[93]: Apple_units.sum()
```

```
[93]: applesoldunits              6348
      appleaccruedpreorderunits      0
      applelumpedpreorderunits       0
      dtype: int64
```

```
[94]: apple_units = pd.DataFrame(data= [6348,0,0],

                     ␣
      ↪index=['applesoldunits','appleaccruedpreorderunits','applelumpedpreorderunits']
                      ,columns=['No of units sold'])
```

```
[95]: apple_units    = apple_units.reset_index()
      apple_units.columns = ['Resources','no of units sold']
```

```
[96]: # total units sold by various types of resources at apple.
      chart = sns.barplot('Resources','no of units sold',data = apple_units)
      chart.set_xticklabels(chart.get_xticklabels(), rotation=45,␣
       ↪horizontalalignment='right')
```

```
[96]: [Text(0, 0, 'applesoldunits'),
       Text(0, 0, 'appleaccruedpreorderunits'),
       Text(0, 0, 'applelumpedpreorderunits')]
```
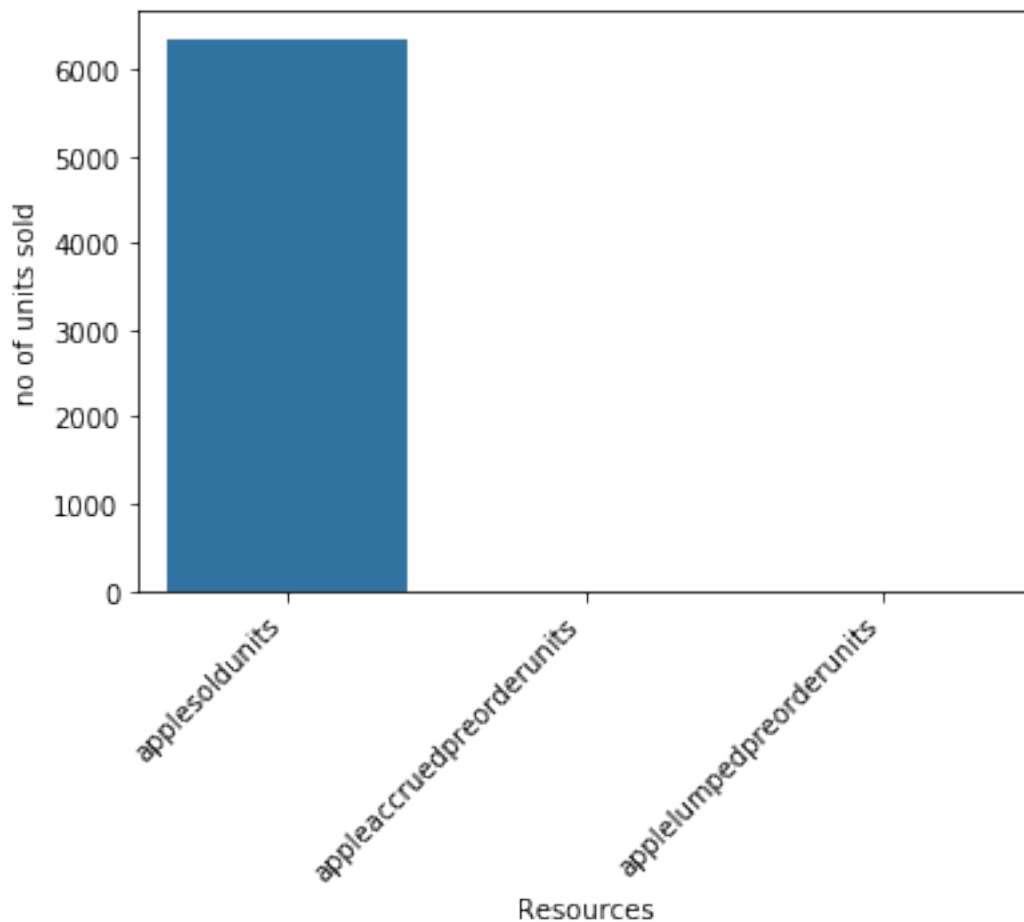

No units were preordered at apple.

```
[97]: revenue_by_format_The_walking_dead =␣
      ↪The_walking_dead[['formattype','totalrevenue']]
```

```
[98]:
```

47

```
revenue_for_print =␣
↪revenue_by_format_The_walking_dead[revenue_by_format_The_walking_dead['formattype']=='print
↪sum()
revenue_for_print
```

[98]:
```
formattype       printprintprintprintprintprintprintprintprintp...
totalrevenue                                          70174946
dtype: object
```

[99]:
```
revenue_for_ebook =␣
↪revenue_by_format_The_walking_dead[revenue_by_format_The_walking_dead['formattype']=='ebook
↪sum()
revenue_for_ebook
```

[99]:
```
formattype       ebookebookebookebookebookebookebookebookebooke...
totalrevenue                                          15669751
dtype: object
```

[100]:
```
revenue_for_audio =␣
↪revenue_by_format_The_walking_dead[revenue_by_format_The_walking_dead['formattype']=='audio
↪sum()
revenue_for_audio
```

[100]:
```
formattype       audioaudioaudioaudioaudioaudioaudioaudioaudioa...
totalrevenue                                          28402266
dtype: object
```

[101]:
```
Revenue_byformat_The_walking_dead = pd.DataFrame(data=␣
↪[70174946,15669751,28402266],
                              ␣
↪index=['revenue_for_print','revenue_for_ebook','revenue_for_audio']
                              ,columns=['Revenue'])
```

[102]:
```
#Revenue by format type for The walking Dead
plt.figure(dpi=100)
explode_list= [0.0,0.05,0.1]
color_list = ['lightblue','brown','green']
Revenue_byformat_The_walking_dead['Revenue'].plot(kind='pie',
                     figsize=(10,6),
                     autopct='(%1.1f%%)',
                      startangle= 60,
                    colors = color_list,
                    pctdistance = 1.58,
                    explode = explode_list)
plt.title('Revenue by formattype Distribution of The Walking Dead')
plt.legend(Revenue_byformat_The_walking_dead.index,loc = 'best')
```

[102]:
```
<matplotlib.legend.Legend at 0x1db51e9dd68>
```

Revenue by formattype Distribution of The Walking Dead

Legend:
- revenue_for_print
- revenue_for_ebook
- revenue_for_audio

(61.4%)
revenue_for_print

(24.9%)
revenue_for_audio

revenue_for_ebook

(13.7%)
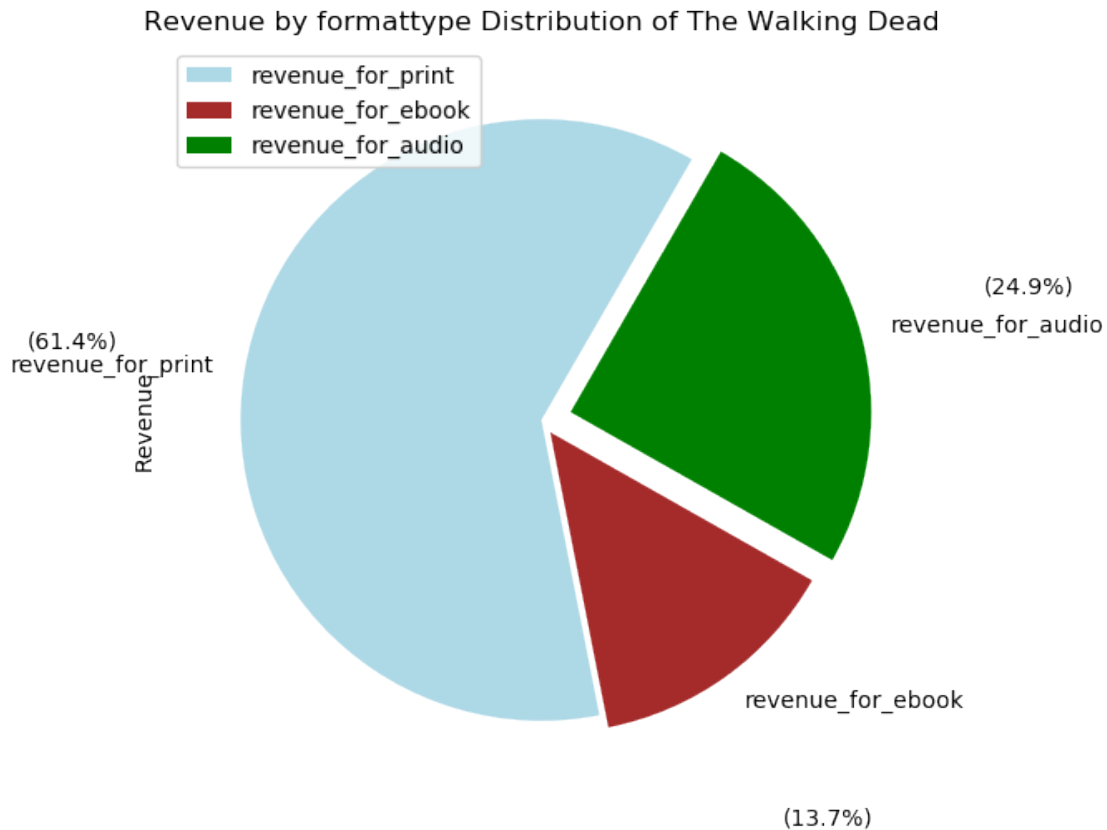
Most of the Revenue of The Walking dead came from print formattype.

```
[103]: revenue_by_format_Alone_in_the_dark=␣
       ↪Alone_in_the_dark[['formattype','totalrevenue']]
```

```
[104]: revenue_for_print =␣
       ↪revenue_by_format_Alone_in_the_dark[revenue_by_format_Alone_in_the_dark['formattype']=='pri
       ↪sum()
       revenue_for_print
```

```
[104]: formattype       printprintprintprintprintprintprintprintprintp...
       totalrevenue                                           2269378
       dtype: object
```

```
[105]: revenue_for_ebook =␣
       ↪revenue_by_format_Alone_in_the_dark[revenue_by_format_Alone_in_the_dark['formattype']=='ebo
       ↪sum()
       revenue_for_ebook
```

```
[105]: formattype       ebookebookebookebookebookebookebookebookebooke...
       totalrevenue                                           5966299
       dtype: object
```

```
[106]: revenue_for_audio =␣
       ↪revenue_by_format_Alone_in_the_dark[revenue_by_format_Alone_in_the_dark['formattype']=='aud
       ↪sum()
       revenue_for_audio
```

```
[106]: formattype      audioaudioaudioaudioaudioaudioaudioaudioaudioa...
       totalrevenue                                          2066812
       dtype: object
```

```
[107]: Revenue_byformat_Alone_in_the_dark = pd.DataFrame(data=␣
       ↪[2269378,5966299,2066812],
                                           ␣
       ↪index=['revenue_for_print','revenue_for_ebook','revenue_for_audio']
                                   ,columns=['Revenue'])
```

```
[108]: #revenue by format type for Alone in the Dark
       plt.figure(dpi=100)
       explode_list= [0.0,0.05,0.1]
       color_list = ['lightblue','brown','green']
       Revenue_byformat_Alone_in_the_dark['Revenue'].plot(kind='pie',
                           figsize=(10,6),
                           autopct='(%1.1f%%)',
                            startangle= 20,
                       colors = color_list,
                       pctdistance = 1.38,
                       explode = explode_list)
       plt.title('Revenue by formattype Distribution of Alone in the dark')
       plt.legend(Revenue_byformat_Alone_in_the_dark.index,loc = 'upper left')
```

```
[108]: <matplotlib.legend.Legend at 0x1db51ef42e8>
```

Revenue by formattype Distribution of Alone in the dark
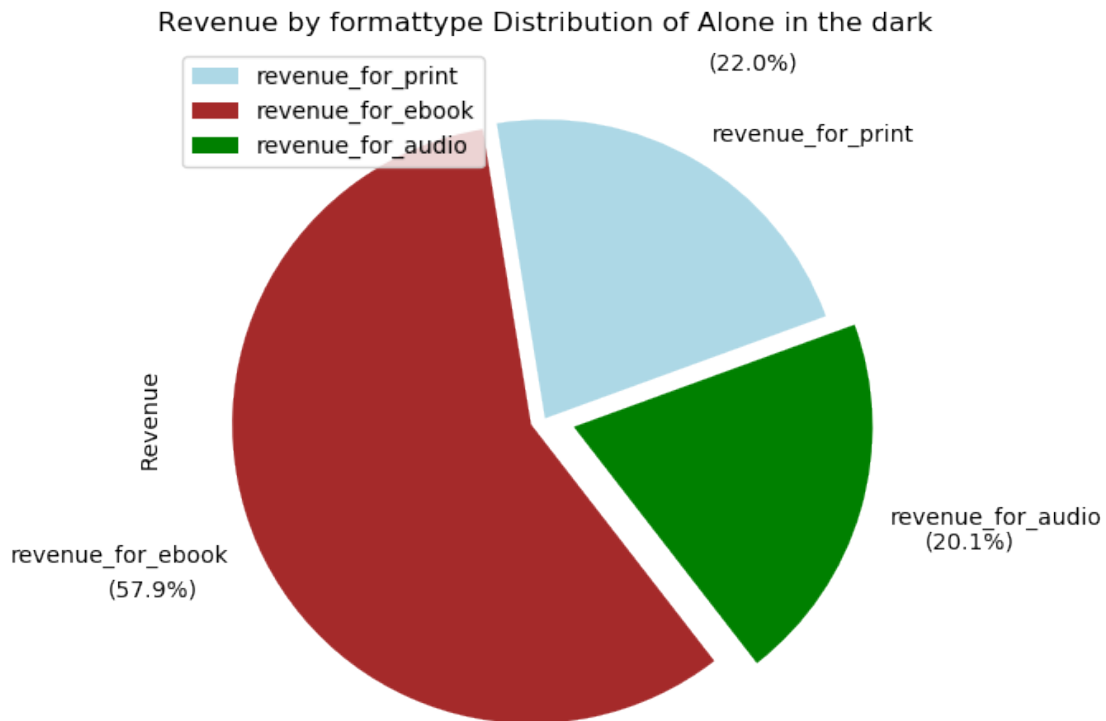
Most of the Revenue of Alone in the Dark came from ebook format type.

- I can forecast the sell of units or revenue generated by the 2 types of entities given in the dataset by their formattype using method such as moving average.Because, I feel that simple method such as moving average would be most ideal for this type of data set.

- However, due to constraint of time I could only perform the analysis part.

[ ]: